



A characterization of the Non-Degenerate Source Condition in Super-Resolution

Vincent Duval

► To cite this version:

Vincent Duval. A characterization of the Non-Degenerate Source Condition in Super-Resolution. Information and Inference, 2020, 9 (1), 10.1093/imaiai/iaz002 . hal-01665805

HAL Id: hal-01665805

<https://inria.hal.science/hal-01665805>

Submitted on 18 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A characterization of the Non-Degenerate Source Condition in Super-Resolution

Vincent Duval^{*†}

December 18, 2017

Abstract

In a recent article, Schiebinger *et al.* provided sufficient conditions for the noiseless recovery of a signal made of M Dirac masses given $2M + 1$ observations of, *e.g.*, its convolution with a Gaussian filter, using the Basis Pursuit for measures. In the present work, we show that a variant of their criterion provides a necessary and sufficient condition for the Non-Degenerate Source Condition (NDSC) which was introduced by Duval and Peyré to ensure support stability in super-resolution. We provide sufficient conditions which, for instance, hold unconditionally for the Laplace kernel provided one has at least $2M$ measurements. For the Gaussian filter, we show that those conditions are fulfilled in two very different configurations: samples which approximate the uniform Lebesgue measure or, more surprisingly, samples which are all confined in a sufficiently small interval.

Keywords: Super-resolution; ℓ^1 norm; Total Variation minimization; LASSO for measures; T-Systems

AMS subject classifications: 90C25, 68U10

1 Introduction

Super-resolution imaging, or observing small structures below the diffraction limit, is a fundamental problem in optical science, concerning applications such as microscopy [32], astronomy [37], medical imaging [22]. It is also of great importance in radar sensing [34] and geophysics [30]. While practical methods have made tremendous progress, the mathematical understanding of that problem is still limited. With the seminal works on sparse super-resolution [11, 5, 7] and the simultaneous emergence of “off-the-grid” methods for line spectral estimation [3, 45], several theoretical advances have been made in the recovery of sparse signals in a continuous domain.

Typically, given a domain $X \subseteq \mathbb{R}$, one wants to recover the signal

$$m_0 \stackrel{\text{def.}}{=} \sum_{i=1}^M a_i \delta_{x_i} \tag{1}$$

²CEREMADE, Université Paris-Dauphine, Place du Marechal De Lattre De Tassigny, 75775 PARIS CEDEX 16, FRANCE.

¹INRIA, Mokaplan

where $\{x_i\}_{i=1}^M \subset X$ are spikes locations and $\{a_i\}_{i=1}^M \subset \mathbb{C}$ are their amplitudes, but one only has access to the observation

$$y = \sum_{i=1}^M a_i \varphi(x_i) + w, \quad (2)$$

where $\varphi(x) \in \mathcal{H}$ is the impulse response, or *point spread function* (in a generalized sense), \mathcal{H} is a Hilbert space, and w is some additive noise. For instance, one may consider $X = [0, 1]$, $\mathcal{H} = \mathbb{C}^K$, and let $\varphi(x) = (c_{j_k}(m))_{1 \leq k \leq K}$ be a vector containing the Fourier coefficients of δ_x for K prescribed frequencies (see [7]). Alternatively, one may choose $\mathcal{H} = L^2(\mathbb{R})$ and $\varphi(x) = \tilde{\psi}(\cdot - x)$ is the impulse response of a linear translation invariant filter [46]. A more realistic setting is to consider that the convolution has been sampled, which can be handled using the more general choice

$$\varphi(x) : s \mapsto \psi(x, s) \quad \text{and} \quad \mathcal{H} = L^2(S, P_S), \quad (3)$$

where ψ encodes the impulse response (e.g. $\psi(x, s) = \tilde{\psi}(s - x)$ for a convolution) and P_S is some (positive) measure (typically, but not necessarily, with finite support) on $S \subseteq \mathbb{R}$, which we call sampling measure.

A ground-breaking result of [7] is that a measure m_0 on $X = [0, 1]$ (with periodic boundary condition) can be recovered from its first $2f_c + 1$ Fourier coefficients, *i.e.* for

$$w = 0, \quad \varphi(x) \stackrel{\text{def.}}{=} \left(e^{-2i\pi kx} \right)_{-f_c \leq k \leq f_c} \quad \text{and} \quad y^{(0)} = \int_X \varphi(x) dm_0(x), \quad (4)$$

by solving the convex minimization problem

$$\min_{m \in \mathcal{M}(X)} |m|(X) \quad \text{s.t.} \quad \int_X \varphi(x) dm(x) = y^{(0)}, \quad (5)$$

provided m_0 is made of spikes which are at least separated by a distance of C/f_c where $C \geq 1.26$ (see [19]). The minimization above is performed in $\mathcal{M}(X)$, the space of all Radon measures on X , and $|m|(X)$ denotes the dual norm of $\|\cdot\|_\infty$ (also known as *total variation*).

That separation condition is in fact a fundamental limit of total variation recovery when looking for spikes with arbitrary sign (or phase), hence the term “super-resolution” in this context is arguable. Indeed, if one observes K measurements, *i.e.* $\varphi(x) = (\varphi_k(x))_{1 \leq k \leq K} \in \mathbb{C}^K$ with φ_k smooth, the optimality of m_0 is equivalent to being able to interpolate its sign/phase with a function of the form

$$\eta(x) = \sum_{k=1}^K \alpha_k \varphi_k(x), \quad \|\eta\|_\infty \leq 1. \quad (6)$$

The span of $\{\varphi_k\}_{1 \leq k \leq K}$ being finite-dimensional, there exists a constant $C > 0$ such that the Bernstein inequality $\|\eta'\|_\infty \leq C \|\eta\|_\infty$ holds, which prevents η to interpolate the values -1 and $+1$ at arbitrarily close locations. More quantitative bounds are provided in [43] for several families of functions, and in [17] for Fourier measurements, explaining that (5) is unable to resolve opposite spikes that are too close to each other.

However, the situation *radically changes if all the amplitudes of m_0 are positive*. In that case, the authors of [11] notice that for impulse responses $\varphi = (\varphi_k)_{1 \leq k \leq K}$ which satisfy some T -system property [29], m_0 is the unique solution to (5) regardless of the separation of its spikes. That phenomenon, which is investigated further in [40], paves

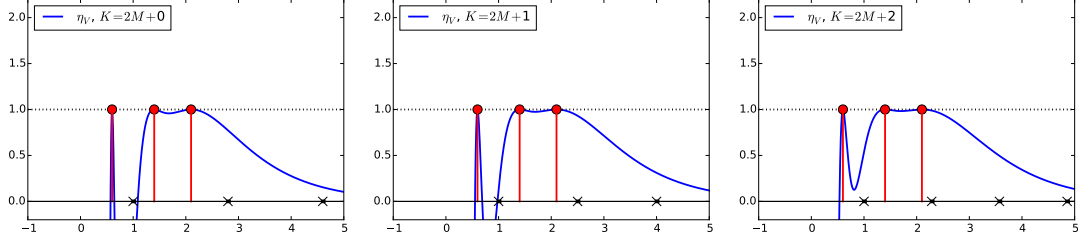


Figure 1: Vanishing derivatives precertificate for Laplace observations ($K = 2M, 2M+1$, and $2M+2$ measurements). It is sufficient to observe $K = 2M$ samples.

the way for *effective superresolution using sparse convex methods*, considering variants which impose nonnegativity such as

$$\min_{m \in \mathcal{M}^+(X)} |m|(X) \quad \text{s.t.} \quad \int_X \varphi(x) dm(x) = y^{(0)}. \quad (7)$$

Nevertheless, a critical issue in super-resolution problems is the stability to noise. The present paper discusses the stability to noise, in the sense of support stability, of such reconstructions for arbitrarily close spikes when considering the regularized inverse problem

$$\min_{m \in \mathcal{M}^+(X)} \lambda |m|(X) + \frac{1}{2} \left\| \int_X \varphi(x) dm(x) - y \right\|_{\mathcal{H}}^2, \quad (8)$$

given some noisy version $y = (y_k)_{1 \leq k \leq K}$ of $y^{(0)} = \int_X \varphi dm_0$. As explained in [17], the crux of the matter is to understand whether the *vanishing derivatives precertificate*, a generalization of the Fuchs precertificate [20], is able to take the value 1 on $\{x_i\}_{i=1}^M$ while being less than 1 in $X \setminus \{x_i\}_{i=1}^M$. We call this property the *Non-Degenerate Source Condition*. Building on the work [40], we propose in Section 3 a necessary and sufficient condition for the Non-Degenerate Source Condition. We show that for some filters, that condition holds for any choice of S and P_S (that is, the repartition of the measurements and their respective weights in the \mathcal{H} -norm) provided there are at least $2M$ samples, regardless of the separation of the points x_1, \dots, x_M .

Such is the case of the Laplace filter (see Figure 1), as shown in Section 4, which provides theoretical foundations for the techniques used in [14] for microscopy imaging. Admittedly, the constructed dual certificates decay slowly when going away from the x_i 's, indicating that the stability constants are not so good, and the problem is numerically unstable at large scales. But reconstructing a measure from a few Laplace measurements is a fundamentally difficult problem, and any reconstruction method is bound to face similar limitations. Our results indicate that (8) is able to provide a sound reconstruction method which is robust to small perturbations, at least for a small number of spikes. This property is confirmed by the numerical experiments of [14].

The case of the convolution with a Gaussian kernel is more subtle. We show in Section 4.2 that (7) and (8) yield a support-stable reconstructions provided that the convolution is fully observed (and P_S is the Lebesgue measure). In general, this property does not hold for any sampling measure P_S , but it holds for discrete measures P_S which approximate the Lebesgue measure sufficiently well, see Figure 2. Surprisingly, this support recovery property also holds in the case where the convolution is sampled in a very small interval. We find this “confined sampling” setting quite counter-intuitive, as the samples can be placed anywhere (provided they are in some small interval) with respect to the spikes to recover, contrary *e.g.* to [2] which requires at least two samples in the neighborhood each spike to recover.

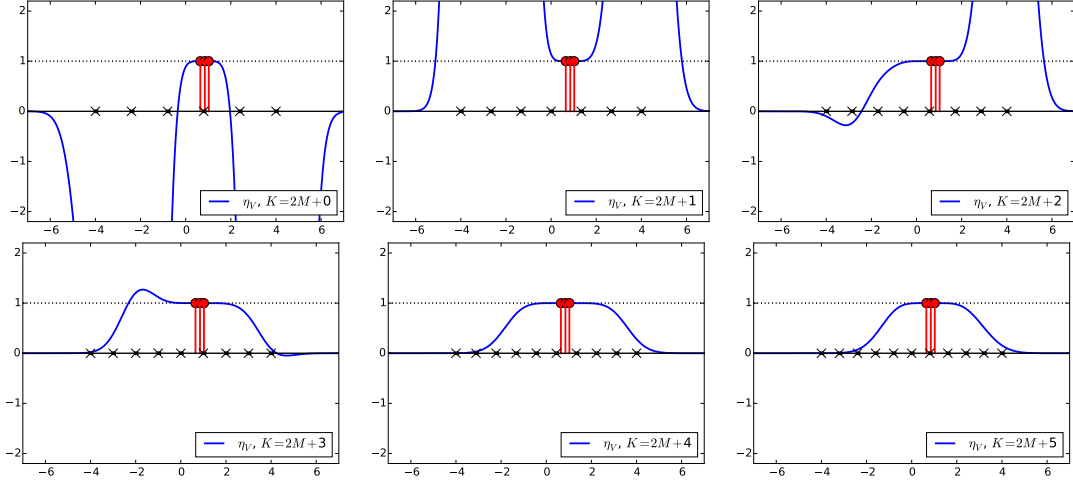


Figure 2: Vanishing derivatives precertificate for the sampled convolution with a Gaussian kernel $e^{-(x-s)^2}$ (the locations of the samples s are indicated by black crosses). The stability criterion ($\eta_V < 1$) is not always satisfied, but if the sampling measure P_S approximates the Lebesgue measure sufficiently well, it does hold.

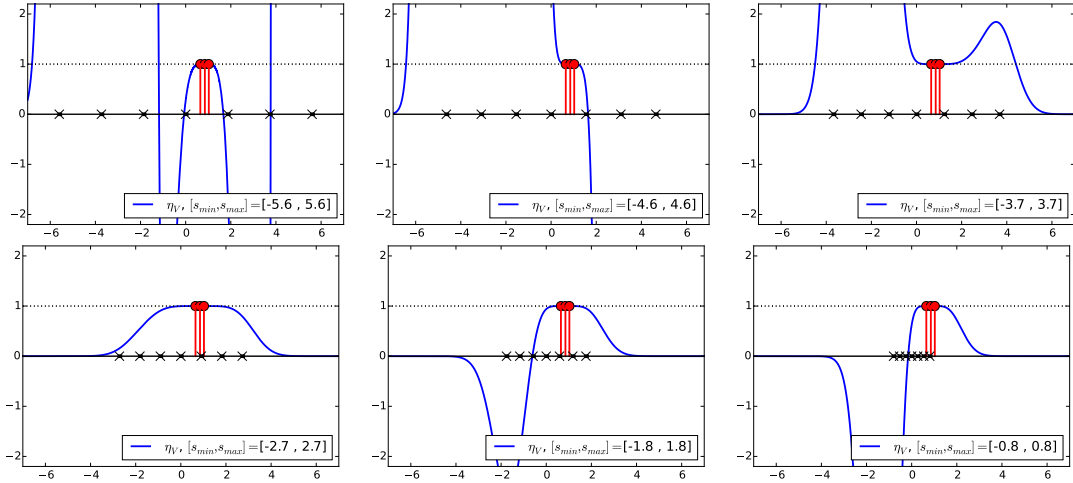


Figure 3: Vanishing derivatives precertificate for the sampled convolution with a Gaussian kernel $e^{-(x-s)^2}$ (the locations of the samples s are indicated by black crosses). If the samples are sufficiently confined (bottom row), the precertificate is valid, $\eta_V < 1$.

1.1 Previous Works

Discrete ℓ^1 regularization The use of ℓ^1 -type techniques for the resolution of inverse problems originates from works in geophysics for seismic exploration [9, 39, 31]. Their popularization in signal processing and statistics comes respectively from the seminal works of Donoho [15, 42] and Tibshirani [47]. One usually defines a discrete grid and solves a variational problem on that grid. The literature concerning the robustness of such models is abundant, let us mention [16, 20, 49] which guarantees for exact support recovery, and [21] which bounds the L^2 recovery error.

Off-the-grid methods However, using a discrete grid only approximates the problem to be solved [44, 24], generally breaking the sparsity assumption of the signal (the so-called *basis mismatch* phenomenon [8]): some resulting artifacts are described in [18]. To circumvent the problem, several researchers [11, 3, 5, 7] proposed to work in a gridless setting, turning the finite dimensional *basis pursuit* or *lasso* problems into variational problems in the space of Radon measures. Within that new framework, the notion of minimum separation of the spikes is the cornerstone of several identifiability results [11, 7, 45] as well as robustness results [6, 26, 17]. The particular case of positive sources, which does not impose any separation of the spikes, has been studied in [11, 41] for identifiability, and in [13] for noise robustness (see also [33] in a discrete setup). In dimension $d \geq 2$, the situation is considerably more difficult, the stability constants are different, as shown in [36].

Non variational methods Let us also note that, while we focus here on a variant of the Lasso for super-resolution, there is a large panel of methods designed to tackle super-resolution problems. To name a few, MUSIC [35] ESPRIT [27], finite rate-of-innovation [4], matrix pencil [25] or Cadzow denoising [10] techniques are also worth considering. Let us also mention [12, 48] which tackle the super-resolution problem using the (non-convex) tools from sparsity (ℓ^0 norm or k -sparse vectors). Note that there is a deep connection between super-resolution and machine learning problems such as the recovery of probability densities from random empirical generalized moments [23], which paves the way for efficient methods in the setting of random measurements. It seems that, for a given problem, the method of choice will strongly depend on the desired properties: flexibility on the measurement operator, robustness to noise, ability to handle signals with signs or phases. We do not claim that the Lasso for measures is universally better than the above-mentioned methods, we simply regard (7) as an interesting variational problem that we wish to understand.

1.2 Contributions

In the present work, we propose a necessary and sufficient condition for the *Non-Degenerate Source Condition* (Theorem 3) which ensures support stability when trying to solve the super-resolution problem using (8). This condition can be seen as some infinitesimal variant of the Determinantal condition in [41], its main advantage is that it ensures the non-vanishing of the second derivatives.

We show that when the observations consist of several samples of an integral transform, it is possible to ensure a priori, for some kernels, that those conditions hold regardless of the positions of the (sufficiently many) samples. Such is the case of the partial Laplace transform or the sampled convolution with a Gaussian (with an L^1 normalization). In the case of the unnormalized Gaussian, we show that similar results hold

provided the samples approximate the uniform Lebesgue measure sufficiently well, or, on the contrary, if they are drawn in a sufficiently small interval.

The proposed results heavily rely on the techniques introduced in [41] in the field of sparse recovery. Nevertheless, we develop here a self-contained theory, which benefits from the following key differences:

- Our approach is free of normalization, whereas [41] requires an L^1 -type normalization of the atoms. Despite the good empirical results reported in [41], it is arguable whether this normalization is the one to use, as the problem $(\mathcal{P}_\lambda(y))$ seems naturally Hilbertian. Also, we clarify the relationship between the L^1 normalization and the invoked conditions.
- On a related topic (see Section 3.3), the approach in [41] naturally requires $2M + 1$ observations of the filtered signal (where m_0 has M spikes), our approach without normalization only needs $2M$, which is the optimal requirement.
- Most importantly, we ensure the stability of the support at low noise, whereas [41] only provides identifiability.

1.3 Notations and Preliminaries

Functional spaces Let X be an interval of \mathbb{R} . We denote by $\mathcal{C}_0(X)$ the space of \mathbb{R} -valued continuous functions which vanish “at infinity”, *i.e.* for all $\varepsilon > 0$, there exists a compact subset $K \subset X$ such that

$$\forall x \in X \setminus K, \quad |\eta(x)| \leq \varepsilon.$$

More explicitly,

- If $X = [a, b] \subset \mathbb{R}$, $\mathcal{C}_0(X) = \mathcal{C}([a, b])$ is the space of continuous functions on $[a, b]$.
- If $X = [a, b) \subset \mathbb{R}$, $\mathcal{C}_0(X)$ is the space of continuous functions η on $[a, b)$ such that $\lim_{x \rightarrow b} \eta(x) = 0$.
- If $X = (a, b) \subseteq \mathbb{R}$, $\mathcal{C}_0(X)$ is the space of continuous functions η on (a, b) such that $\lim_{x \rightarrow a} \eta(x) = \lim_{x \rightarrow b} \eta(x) = 0$.

Endowed with the supremum norm, $\mathcal{C}_0(X)$ is a Banach space whose dual is the space $\mathcal{M}_b(X)$ of finite Radon measures (see [1]). We denote by $\mathcal{M}_b^+(X)$ the set of nonnegative Radon measures. In the rest of the paper, we regard $\mathcal{C}_0(X)$ and $\mathcal{M}_b(X)$ as locally convex topological vector spaces, and we endow them respectively with their weak and weak-* topologies. With that choice, both spaces are dual to each other.

The functional

$$G(m) \stackrel{\text{def.}}{=} m(X) + \chi_{\mathcal{M}_b^+(X)} = \begin{cases} m(X) & \text{if } m \text{ is nonnegative,} \\ +\infty & \text{otherwise.} \end{cases}$$

can be written as a support function

$$G(m) = \sup \left\{ \int_X \eta dm : \eta \in \mathcal{C}_0(X) \text{ and } \sup_{x \in X} \eta(x) \leq 1 \right\},$$

so that its subdifferential is

$$\partial G(m) = \left\{ \eta \in \mathcal{C}_0(X) : \sup \eta \leq 1 \text{ and } \int_X \eta dm = m(X) \right\}. \quad (9)$$

Linear operators Let \mathcal{H} be a real separable Hilbert space. Let $\varphi \in \mathcal{C}(X; \mathcal{H})$ be a continuous \mathcal{H} -valued function such that for all $p \in \mathcal{H}$, $(x \mapsto \langle p, \varphi(x) \rangle_{\mathcal{H}}) \in \mathcal{C}_0(X)$. Note that by the Banach-Steinhaus theorem, $\sup_{x \in X} \|\varphi(x)\|_{\mathcal{H}} < +\infty$.

Then, the linear operator $\Phi : \mathcal{M}_b(X) \rightarrow \mathcal{H}$,

$$\Phi m \stackrel{\text{def.}}{=} \int_X \varphi dm \quad (10)$$

(where the above quantity is a Bochner integral [50, Sec. V.5]) is weak-* to weak continuous, as shown by the equality

$$\langle \Phi m, p \rangle_{\mathcal{H}} = \left\langle \int_X \varphi dm, p \right\rangle_{\mathcal{H}} = \int_X \langle \varphi(x), p \rangle_{\mathcal{H}} dm(x).$$

It also shows that the adjoint $\Phi^* : \mathcal{H} \rightarrow \mathcal{C}_0(X)$ has full domain, with $(\Phi^* p) : x \mapsto \langle \varphi(x), p \rangle_{\mathcal{H}}$. We note that Φ^* is continuous (from the strong to the strong topology as well as) from the weak to the weak topology.

Admissible Kernels We say φ is in $\text{KER}^{(k)}$ if $\varphi \in \mathcal{C}^k(X; \mathcal{H})$ and

- for all $p \in \mathcal{H}$, $(x \mapsto \langle p, \varphi(x) \rangle_{\mathcal{H}}) \in \mathcal{C}_0(X)$.
- for all $0 \leq i \leq k$, $\sup_{x \in X} \|\varphi^{(i)}(x)\|_{\mathcal{H}} < +\infty$,

Given $\mathbf{x} = (x_1, \dots, x_M) \in X^M$, we denote by $\Phi_{\mathbf{x}} : \mathbb{R}^M \rightarrow \mathcal{H}$ the linear operator such that

$$\forall a \in \mathbb{R}^M, \quad \Phi_{\mathbf{x}}(a) \stackrel{\text{def.}}{=} \sum_{i=1}^M a_i \varphi(x_i),$$

and by $\Gamma_{\mathbf{x}} : \mathbb{R}^{2M} \rightarrow \mathcal{H}$ the linear operator defined by

$$\Gamma_{\mathbf{x}} b \stackrel{\text{def.}}{=} \sum_{i=1}^M (b_{2i-1} \varphi(x_i) + b_{2i} \varphi'(x_i)). \quad (11)$$

We adopt the following matricial notation

$$\Phi_{\mathbf{x}} = (\varphi(x_1) \ \dots \ \varphi(x_M)) \quad \text{and} \quad \Gamma_{\mathbf{x}} = (\varphi(x_1), \varphi'(x_1) \ \dots \ \varphi(x_M), \varphi'(x_M)),$$

where the “columns” of those “matrices” are elements of \mathcal{H} . In particular, the adjoint operator $\Phi_{\mathbf{x}}^* : \mathcal{H} \rightarrow \mathbb{R}^M$ is given by

$$\forall p \in \mathcal{H}, \quad \Phi_{\mathbf{x}}^* p = ((\Phi^* p)(x_i))_{1 \leq i \leq M}.$$

Given $x_0 \in X$, we denote by $\varphi_k \in \mathcal{H}$ the k^{th} derivative of φ at x_0 , i.e.

$$\varphi_k \stackrel{\text{def.}}{=} \varphi^{(k)}(x_0). \quad (12)$$

In particular, $\varphi_0 = \varphi(x_0)$.

Given $k \in \mathbb{N}$, we define

$$\Psi_k \stackrel{\text{def.}}{=} (\varphi_0 \ \varphi_1 \ \dots \ \varphi_k). \quad (13)$$

If $\Psi_k : \mathbb{R}^{k+1} \rightarrow \mathcal{H}$ has full column rank, we define its pseudo-inverse as $\Psi_k^+ \stackrel{\text{def.}}{=} (\Psi_k^* \Psi_k)^{-1} \Psi_k^*$. Similarly, we denote $\Gamma_{\mathbf{x}}^+ \stackrel{\text{def.}}{=} (\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})^{-1} \Gamma_{\mathbf{x}}^*$ provided $\Gamma_{\mathbf{x}}$ has full column rank.

2 The Blasso with nonnegativity constraints

In this section, we describe the main properties of the positive Beurling LASSO (BLASSO⁺) that are needed in the paper. These are mainly minor adaptations of the properties stated in [17, 13] for the BLASSO, and we state them here without proof.

We assume that we are given a Hilbert space \mathcal{H} and some kernel $\varphi \in \text{KER}^{(k)}$ for some $k \geq 2$ as described above. The choice of φ and \mathcal{H} is discussed in detail in Section 2.4

2.1 Variational Problems in the Space of Measures

Our goal is to recover an unknown input measure

$$m_0 = m_{a,\mathbf{x}} \stackrel{\text{def.}}{=} \sum_{i=1}^M a_i \delta_{x_i}$$

with $(a_i, x_i) \in \mathbb{R} \times X$ from the observations $y = \Phi m_0 + w \in \mathcal{H}$, where Φ is the linear operator defined in (10), and $w \in \mathcal{H}$ is some additive noise. In this paper, we assume that the amplitudes a_i are positive.

To perform the reconstruction, we solve the following variant of the BLASSO, adding a nonnegativity constraint, that we denote by BLASSO⁺,

$$\min_{m \in \mathcal{M}_b^+(X)} \lambda m(X) + \frac{1}{2} \|\Phi m - y\|_{\mathcal{H}}^2, \quad (\mathcal{P}_\lambda(y))$$

where $\lambda > 0$ is some regularization parameter. In the noiseless case ($w = 0$), one might as well solve the constrained problem

$$\min_{\substack{m \in \mathcal{M}_b^+(X) \\ \Phi m = \Phi m_0}} m(X). \quad (\mathcal{P}_0(\Phi m_0))$$

It worth noting that $(\mathcal{P}_0(\Phi m_0))$ is the limit of $(\mathcal{P}_\lambda(y))$ as $\lambda \rightarrow 0$ and $\|w\|_{\mathcal{H}} \rightarrow 0$ in a suitable way (see [17] for the case of the BLASSO).

2.2 Low noise behavior and exact support recovery

From now on, in order to simplify the discussion, we assume that $x_i \in \text{int}(X)$ for all $1 \leq i \leq M$, i.e. $\inf X < x_i < \sup X$.

Exact support recovery In [17], we have provided an almost sharp sufficient condition for the exact support recovery at low noise. A (simplified) version of that condition is the following. If $\Gamma_{\mathbf{x}}$ (see (11)) has full rank, we define the vanishing derivatives pre-certificate as

$$\eta_V \stackrel{\text{def.}}{=} \Phi^* p_V, \quad \text{where} \quad (14)$$

$$p_V \stackrel{\text{def.}}{=} \text{argmin} \{ \|p\|_{\mathcal{H}} : p \in \mathcal{H}, (\Phi^* p)(x_i) = 1, (\Phi^* p)'(x_i) = 0 \text{ for all } 1 \leq i \leq M \}. \quad (15)$$

In other words, $p_V = \Gamma_x^{+,*} u$ where $u \stackrel{\text{def.}}{=} (1, 0, 1, 0, \dots, 1, 0)^T \in \mathbb{R}^{2M}$, and $\Gamma_x^{+,*} = \Gamma_x (\Gamma_x^* \Gamma_x)^{-1}$ is the Moore-Penrose pseudo-inverse of Γ_x . We note that by optimality of p_V in (14), there exists some coefficients $\{\alpha_i, \beta_i\}_{1 \leq i \leq M} \subset \mathbb{R}$ such that $p_V = \sum_{i=1}^M (\alpha_i \varphi(x_i) + \beta_i \varphi'(x_i))$, and η_V is the unique function of the form

$$\eta_V(x) = \sum_{i=1}^M (\alpha_i \langle \varphi(x), \varphi(x_i) \rangle_{\mathcal{H}} + \beta_i \langle \varphi'(x), \varphi'(x_i) \rangle_{\mathcal{H}}) \quad (16)$$

which satisfies $\eta_V(x_i) = 1$ and $\eta_V'(x_i) = 0$ for all $i \in \{1, \dots, M\}$.

Definition 1 (NDSC). Assume that $\Gamma_{\mathbf{x}}$ has full rank. We say that m_0 satisfies the *Non Degenerate Source Condition* if $\eta_V(x) < 1$ for all $x \in X \setminus \{x_1, \dots, x_M\}$ and $\eta_V''(x_i) < 0$ for all $1 \leq i \leq M$.

As the next result shows, the nondegeneracy of η_V almost characterizes the support stability of the BLASSO⁺ at low noise. When it holds, it entails $|\tilde{a}_i(\lambda, w) - a_i| = O(\|w\|_{\mathcal{H}})$ and $|\tilde{x}_i(\lambda, w) - x_i| = O(\|w\|_{\mathcal{H}})$ for $\lambda = \frac{1}{\alpha} \|w\|_{\mathcal{H}}$.

Theorem 1 ([17]). Assume that $\Gamma_{\mathbf{x}}$ has full rank and that m_0 satisfies the *Non Degenerate Source Condition*. Then there exists $\alpha > 0$, $\lambda_0 > 0$ such that for all $0 \leq \lambda \leq \lambda_0$ and $\|w\|_{\mathcal{H}} \leq \alpha\lambda$, the solution $\tilde{m}_{\lambda, w}$ to $(\mathcal{P}_{\lambda}(y))$ is unique and is composed of exactly M spikes, with $\tilde{m}_{0,0} = m_0$.

Moreover one may write $\tilde{m}_{\lambda, w} = \sum_{i=0}^M \tilde{a}_i(\lambda, w) \delta_{\tilde{x}_i(\lambda, w)}$, where \tilde{a}_i and \tilde{x}_i are the restriction of \mathcal{C}^1 functions on a neighborhood of $(0, 0)$.

Conversely, if $\Gamma_{\mathbf{x}}$ has full rank and the solutions can be written in the above form with functions \tilde{a}_i and \tilde{x}_i which are \mathcal{C}^1 and satisfy $\tilde{a}_i(0, 0) = a_i$, $\tilde{x}_i(0, 0) = x_i$, then $\eta_V \leq 1$.

Unfortunately, the Non Degenerate Source Condition (NDSC) is not trivial to ensure a priori. One reason is that it should be checked for any configuration $x_1 < \dots < x_M$. Let us mention [46] which proves that the (NDSC) holds for the Gaussian or Cauchy kernels and signed measures, provided the spikes are sufficiently separated. As noted above in the introduction, in the signed case, the (NDSC) generally cannot hold if spikes with opposite signs are too close (see [17, Corollary 1] and [43]).

The situation is more favorable in the setting of nonnegative measures that we consider. In Section 3, building upon [41], we show that the (NDSC) holds for specific choices of φ which include the Laplace kernel.

The case of clustered spikes Another approach for nonnegative measures, taken in [13], is to consider spikes which are located in a small neighborhood of a single point $x_0 \in \text{int}(X)$, say $x_i = x_0 + tz_i$, for some $z_i \in \mathbb{R}$ and some small $t > 0$. Let us write $m_t = \sum_{i=1}^M a_i \delta_{x_0 + tz_i}$ and $y_t = \Phi m_t + w$. One may prove that, as $t \rightarrow 0^+$, the corresponding p_V converges (strongly in \mathcal{H}) towards

$$p_W = \operatorname{argmin} \left\{ \|p_W\|_{\mathcal{H}} : (\Phi^* p_W)(x_0) = 1, (\Phi^* p_W)'(x_0) = \dots = (\Phi^* p_W)^{(2M-1)}(x_0) = 0 \right\} \quad (17)$$

provided $\varphi \in \text{KER}^{(2M-1)}$ and $(\varphi_0, \dots, \varphi_{2M-1})$ has full rank. As a result $\eta_V = \Phi^* p_V$ converges (uniformly on X) towards $\eta_W \stackrel{\text{def.}}{=} \Phi^* p_W$, and the same holds for the derivatives up to the regularity of φ . Letting $\Psi_{2M-1} \stackrel{\text{def.}}{=} (\varphi_0, \dots, \varphi_{2M-1})$, we note as above that $p_W \stackrel{\text{def.}}{=} \Psi_{2M-1}^{+,*} v$ where $v \stackrel{\text{def.}}{=} (1, 0, \dots, 0)^T \in \mathbb{R}^{2M}$. Moreover, η_W is the unique function of the form

$$\eta_W : x \mapsto \sum_{k=0}^{2M-1} \beta_k \left\langle \varphi(x), \varphi^{(k)}(x_0) \right\rangle_{\mathcal{H}},$$

which satisfies

$$\eta_W(x_0) = 1, \eta_W'(x_0) = 0, \dots, \eta_W^{(2M-1)}(x_0) = 0. \quad (18)$$

Definition 2 (($2M - 1$)-Nondegeneracy). We say that a point x_0 is ($2M - 1$) *Non Degenerate* if

- $\varphi \in \text{KER}^{(2M)}$,
- $\Psi_{2M-1} \stackrel{\text{def.}}{=} (\varphi_0, \dots, \varphi_{2M-1})$ has full rank,

- $\eta_W^{(2M)}(x_0) < 0$,
- for all $x \in X \setminus \{x_0\}$, $\eta_W(x_0) < 1$.

The $(2M-1)$ -Nondegeneracy of a point entails the support stability of finite measures which are clustered around x_0 .

Theorem 2 ([13]). *Assume that $\varphi \in \text{KER}^{(2M+1)}$ and that the point is $(2M-1)$ Non Degenerate. Then there exists $t_0 > 0$, such that m_t satisfies the (NDSC) for for all $0 < t < t_0$.*

More precisely, there exists $\alpha > 0$, $C_R > 0$ and $C > 0$ such that for $0 < t < t_0$ and all $0 < \lambda + \|w\|_{\mathcal{H}} < C_R t^{2M-1}$ with $\|w\|_{\mathcal{H}} \leq \alpha \lambda$,

- *the problem $\mathcal{P}_\lambda(y_t)$ has a unique solution of the form $\sum_{i=1}^M \tilde{a}_i(\lambda, w, t) \delta_{x_0 + t\tilde{z}_i(\lambda, w, t)}$,*
- *the mapping $g_t : (\lambda, w) \mapsto (\tilde{a}_i, \tilde{z}_i)$ is the restriction of a \mathcal{C}^{2M} function,*
- *the following inequality holds*

$$|(\tilde{a}, \tilde{z}) - (a, z)|_\infty \leq C \left(\frac{|\lambda|}{t^{2M-1}} + \frac{\|w\|_{\mathcal{H}}}{t^{2M-1}} \right).$$

Theorem 2 is more quantitative than Theorem 1, as it gives the scaling of the constants with respect to the spacing t of the spikes, describing how it becomes ill-conditioned.

2.3 Ensuring negative second derivatives

In general, the conditions $\eta_V''(x_i) < 0$ and $\eta_W^{(2M)}(x_0) < 0$ in the non degeneracy hypotheses are difficult to ensure a priori. They can be checked numerically, or, in the case of η_W , analytically on a closed form expression. But it is sometimes possible to ensure them a priori: in [13, Proposition 5], we have proved that $\eta_W^{(2M)}(x_0) < 0$ for fully sampled convolutions on translation-invariant domains, under a mild assumption.

In the present paper we propose the following two new Lemmas, which work for any sampling of the impulse response. They follow directly from (16) and the characterization of η_V , in the spirit of Cramer's rule.

Lemma 1. *Assume that $\Gamma_{\mathbf{x}}$ has full rank. Then for all $i \in \{1, \dots, M\}$,*

$$\begin{vmatrix} \langle \varphi(x_1), \varphi(x_1) \rangle_{\mathcal{H}} & \langle \varphi(x_1), \varphi'(x_1) \rangle_{\mathcal{H}} & \cdots & \langle \varphi(x_1), \varphi'(x_M) \rangle_{\mathcal{H}} & 1 \\ \langle \varphi'(x_1), \varphi(x_1) \rangle_{\mathcal{H}} & \langle \varphi'(x_1), \varphi'(x_1) \rangle_{\mathcal{H}} & \cdots & \langle \varphi'(x_1), \varphi'(x_M) \rangle_{\mathcal{H}} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \langle \varphi(x_M), \varphi(x_1) \rangle_{\mathcal{H}} & \langle \varphi(x_M), \varphi'(x_1) \rangle_{\mathcal{H}} & \cdots & \langle \varphi(x_M), \varphi'(x_M) \rangle_{\mathcal{H}} & 1 \\ \langle \varphi'(x_M), \varphi(x_1) \rangle_{\mathcal{H}} & \langle \varphi'(x_M), \varphi'(x_1) \rangle_{\mathcal{H}} & \cdots & \langle \varphi'(x_M), \varphi'(x_M) \rangle_{\mathcal{H}} & 0 \\ \langle \varphi''(x_i), \varphi(x_1) \rangle_{\mathcal{H}} & \langle \varphi''(x_i), \varphi'(x_1) \rangle_{\mathcal{H}} & \cdots & \langle \varphi''(x_i), \varphi'(x_M) \rangle_{\mathcal{H}} & 0 \end{vmatrix} = -\eta_V''(x_i) \det(\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}}). \quad (19)$$

Ensuring that the left hand side is positive provides the condition $\eta_V''(x_M) < 0$, and a similar argument can be derived for the other x_i 's.

The analogous property for η_W is the following.

Lemma 2. *Assume that Ψ_{2M-1} has full column rank.*

Then,

$$\begin{vmatrix} \langle \varphi_0, \varphi_0 \rangle_{\mathcal{H}} & \langle \varphi_0, \varphi_1 \rangle_{\mathcal{H}} & \cdots & \langle \varphi_0, \varphi_{2M-1} \rangle_{\mathcal{H}} & 1 \\ \langle \varphi_1, \varphi_0 \rangle_{\mathcal{H}} & \langle \varphi_1, \varphi_1 \rangle_{\mathcal{H}} & \cdots & \langle \varphi_1, \varphi_{2M-1} \rangle_{\mathcal{H}} & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \langle \varphi_{2M}, \varphi_0 \rangle_{\mathcal{H}} & \langle \varphi_{2M}, \varphi_1 \rangle_{\mathcal{H}} & \cdots & \langle \varphi_{2M}, \varphi_{2M-1} \rangle_{\mathcal{H}} & 0 \end{vmatrix} = -\eta_W^{(2M)}(x_0) \det(\Psi_{2M-1}^* \Psi_{2M-1}). \quad (20)$$

2.4 Reconstruction frameworks

Now, we discuss the choice of the Hilbert space and kernel φ .

Remark 1. Let us note that $(\mathcal{P}_\lambda(y))$ may be written in terms of the autocorrelation function

$$\forall (x, x') \in X^2, \quad R_\varphi(x, x') \stackrel{\text{def.}}{=} \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}.$$

Introducing the autocorrelation operator

$$\begin{aligned} \mathbf{R}_\varphi &\stackrel{\text{def.}}{=} \Phi^* \Phi: \mathcal{M}_b(X) \rightarrow \mathcal{C}_0(X) \\ m &\longmapsto \left(\int_X R_\varphi(\cdot, x') dm(x') \right), \end{aligned}$$

we note that $(\mathcal{P}_\lambda(y))$ is equivalent to

$$\frac{1}{2} \|y\|_{\mathcal{H}}^2 + \min_{m \in \mathcal{M}_b^+(X)} \left(\lambda m(X) + \frac{1}{2} \langle \mathbf{R}_\varphi m, m \rangle - \langle \Phi^* y, m \rangle \right) \quad (\tilde{\mathcal{P}}y) \quad (20)$$

(here $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $\mathcal{C}_0(X)$ and $\mathcal{M}_b(X)$). As a result, everything described above (in particular expression for certificates) can be written in term of the autocorrelation function and its derivatives. Moreover, choosing φ and \mathcal{H} is equivalent (but less intuitive) to choosing the autocorrelation function R_φ .

A typical choice of Hilbert space covered by our framework is the following. Given some positive measure P_S on some interval $S \subseteq \mathbb{R}$, the impulse response is

$$\varphi(x) : s \mapsto \psi(x, s)$$

where ψ is some kernel and $\mathcal{H} = L^2(S, P_S)$. We say that (S, P_S) define a reconstruction framework.

For instance, if $P_S = \sum_{k=1}^K \delta_{s_k}$ and $\psi(x, s) = e^{-(x-s)^2}$, our observation is the sampled convolution of m_0 with the Gaussian kernel,

$$(\Phi m_0)_k = \int_X e^{-(x-s_k)} dm_0(x),$$

and the norm of \mathcal{H} is the standard Euclidean norm in \mathbb{R}^K .

In the following, we usually assume that ψ is smooth in its first argument, and that it is possible to differentiate under the sum symbol,

$$\varphi^{(k)}(x) = (\partial_1)^k \psi(x, \cdot) \in L^2(S, P_S). \quad (21)$$

3 A characterization of the Non-Degenerate Source Condition

In this section we provide a simple criterion which ensures that a measure m_0 is the unique solution to the basis pursuit for measures $(\mathcal{P}_0(\Phi m_0))$, and that its recovery using $(\mathcal{P}_\lambda(y))$ is stable. This criterion is an "infinitesimal version" of the criterion proposed in [41], except that we do not impose any normalisation of the atoms (nor any weight on the total variation). Throughout this section we let $\{x_i\}_{i=1}^M \subset \text{int}(X)$ with $x_1 < \dots < x_M$, and $x_0 \in \text{int}(X)$.

3.1 Rescaled determinants

The following elementary result is a specialization of a standard trick in the study of *extended T-systems* [29]. For sufficiently regular functions $\{u_0, \dots, u_{2M}\}$, define D_V and D_W by

$$\forall t \in X \setminus \{x_i\}_{i=1}^M, \quad D_V(t) \stackrel{\text{def.}}{=} \frac{2}{\prod_{i=1}^M (t - x_i)^2} \begin{vmatrix} u_0(t) & u_1(t) & \cdots & u_{2M}(t) \\ u_0(x_1) & u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_0(x_1) & u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & \vdots & & \vdots \\ u_0(x_M) & u_1(x_M) & \cdots & u_{2M}(x_M) \\ u'_0(x_M) & u'_1(x_M) & \cdots & u'_{2M}(x_M) \end{vmatrix}, \quad (22)$$

$$\forall t \in X \setminus \{x_0\}, \quad D_W(t) \stackrel{\text{def.}}{=} \frac{(2M)!}{(t - x_0)^{2M}} \begin{vmatrix} u_0(t) & u_1(t) & \cdots & u_{2M}(t) \\ u_0(x_1) & u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_0(x_1) & u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & \vdots & & \vdots \\ u_0^{(2M-1)}(x_M) & u_1^{(2M-1)}(x_M) & \cdots & u_{2M}^{(2M-1)}(x_M) \end{vmatrix}. \quad (23)$$

Lemma 3. *If $\{u_0, \dots, u_{2M}\} \subset \mathcal{C}^2(X)$, then D_V has a continuous extension to X , with*

$$D_V(x_i) \stackrel{\text{def.}}{=} \begin{vmatrix} u_0(x_1) & u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_0(x_1) & u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & \vdots & & \vdots \\ u_0(x_i) & u_1(x_i) & \cdots & u_{2M}(x_i) \\ u'_0(x_i) & u'_1(x_i) & \cdots & u'_{2M}(x_i) \\ u''_0(x_i) & u''_1(x_i) & \cdots & u''_{2M}(x_i) \\ \vdots & \vdots & & \vdots \\ u_0(x_M) & u_1(x_M) & \cdots & u_{2M}(x_M) \\ u'_0(x_M) & u'_1(x_M) & \cdots & u'_{2M}(x_M) \end{vmatrix}. \quad (24)$$

Similarly, if $\{u_0, \dots, u_{2M}\} \subset \mathcal{C}^{2M}(X)$, then E has a continuous extension to X , with

$$D_W(x_0) \stackrel{\text{def.}}{=} \begin{vmatrix} u_0(x_1) & u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_0(x_1) & u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & \vdots & & \vdots \\ u_0^{(2M)}(x_M) & u_1^{(2M)}(x_M) & \cdots & u_{2M}^{(2M)}(x_M) \end{vmatrix}. \quad (25)$$

Proof. This is a simple consequence of the Taylor expansions:

$$u_j(t) = u_j(x_i) + u'_j(x_i)(t - x_i) + u''_j(x_i) \frac{(t - x_i)^2}{2} + o((t - x_i)^2),$$

$$u_j(t) = u_j(x_0) + u'_j(x_0)(t - x_0) + \cdots + u_j^{(2M)}(x_0) \frac{(t - x_0)^{2M}}{(2M)!} + o((t - x_0)^{2M}),$$

as $t \rightarrow x_i$ (resp. x_0). □

3.2 Main theorem

As explained in Section 2, the low noise stability of the support of solutions is governed by the *vanishing derivatives precertificate*

$$\eta_V = \Phi^* p_V \quad \text{where} \quad p_V \stackrel{\text{def.}}{=} \operatorname{argmin} \{ \|p\|_{\mathcal{H}} : \Phi^* p(x_i) = 1, \quad (\Phi^* p)'(x_i) = 0 \},$$

provided $\Gamma_{\mathbf{x}}$ has full rank. From (16), we know that η_V can be written as

$$\eta_V(x) = \sum_{i=1}^M (\alpha_i R_\varphi(x, x_i) + \beta_i \partial_2 R_\varphi(x, x_i)), \quad (26)$$

where $R_\varphi(x, y) \stackrel{\text{def.}}{=} \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$ is the autocorrelation function of φ .

In the following, we let

$$(u_0, u_1, u_2, \dots, u_{2M-1}, u_{2M}) \stackrel{\text{def.}}{=} (1, R_\varphi(\cdot, x_1), \partial_2 R_\varphi(\cdot, x_1), \dots, R_\varphi(\cdot, x_M), \partial_2 R_\varphi(\cdot, x_M)), \quad (27)$$

and we define D_V (resp. D_W) as in (22) (resp. (23)).

Theorem 3. Assume that $\varphi \in \text{KER}^{(2)}$ and let $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $a_i > 0$ for all $i \in \{1, \dots, M\}$. Assume moreover that $\Gamma_{\mathbf{x}}$ has full rank.

Then the Non-Degenerate Source condition (NDSC) holds for m_0 if and only if

$$\forall t \in X, \quad D_V(t) > 0. \quad (28)$$

Similarly, let $\varphi \in \text{KER}^{(2M)}$ and assume that Ψ_{2M-1} has full rank.

Then, the $(2M - 1)$ Non Degenerate Source Condition holds at x_0 if and only if

$$\forall t \in X, \quad D_W(t) > 0. \quad (29)$$

The above conditions are clearly reminiscent of the DETERMINANTAL condition in [41]. Three differences should be noted. First, it is an infinitesimal condition: we do not consider neighborhoods of the x_i 's, but instead we consider derivatives. That both makes the proof simpler and provides us with the non-vanishing of the second derivatives of η_V . Second of all, the sign matters: we impose the positivity, which implies that the constructed function η_V is below 1 (whereas the condition in [41] only states that either η_V or $1 - \eta_V$ is below 1). Last, we do not introduce any normalization, and we choose $u_0 = 1$.

Proof. First, we recall that if $\Gamma_{\mathbf{x}}$ has full rank, η_V is the unique function of the form (26) such that $\eta_V(x_i) = 1$, $\eta'_V(x_i) = 0$.

Note that by Lemma 2 and Lemma 3, we know that $\eta''_V(x_i) < 0$ if and only if $D_V(x_i) > 0$. It remains to show that for $t \in X \setminus \{x_i\}_{i=1}^M$, $D_V(t) \neq 0$ if and only if $\eta_V(t) \neq 1$.

Suppose that $D_V(t) \neq 0$ and assume by contradiction that $\eta_V(t) = 1$. Hence the vector $(-1, \alpha_1, \beta_1, \dots, \alpha_M, \beta_M)$ is a non-trivial kernel element of the matrix in (22), which contradicts $D_V(t) \neq 0$. Conversely, if $\eta_V(t) \neq 1$, assume by contradiction that $D_V(t) = 0$. There exists a non trivial vector $(\tilde{\alpha}_0, \tilde{\alpha}_1, \tilde{\beta}_1, \dots, \tilde{\alpha}_M, \tilde{\beta}_M)$ in the kernel of the matrix in (22). If $\tilde{\alpha}_0 = 0$, this yields a non trivial vector in the kernel of $\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}}$, contradicting the fact that $\Gamma_{\mathbf{x}}$ has full rank. Hence, we may assume that $\tilde{\alpha}_0 = -1$, hence the function

$$x \mapsto \sum_{i=1}^M \left(\tilde{\alpha}_i R_\varphi(x, x_i) + \tilde{\beta}_i \partial_2 R_\varphi(x, x_i) \right)$$

is equal to η_V , and $(-1, \tilde{\alpha}_1, \tilde{\beta}_1, \dots, \tilde{\alpha}_M, \tilde{\beta}_M)$ being in the kernel of (22) is equivalent to $\eta_V(t) = 1$, a contradiction.

To conclude, if $D_V(t) > 0$ for all $t \in X$, then by continuity $1 - \eta_V$ cannot vanish in any interval $(\inf X, x_1)$, (x_i, x_{i+1}) or $(x_M, \sup X)$, and $\eta''_V(x_i) < 0$ for $1 \leq i \leq M$.

As a result $1 - \eta_V(t) < 1$ for all $t \in X \setminus \{x_i\}_{i=1}^M$ and the (NDSC) holds. Conversely, if $1 - \eta_V(t) < 1$ for all $t \in X \setminus \{x_i\}_{i=1}^M$ and $\eta_V''(x_i) < 0$, we obtain that $D_V(t) > 0$ for $1 \leq i \leq M$.

Thus, the equivalence is proved in the case of η_V and D_V . The proof for η_W and η_W is the same, *mutatis mutandis*, by observing that if Ψ_{2M-1} has full rank, η_W is the unique function of the form

$$x \mapsto \sum_{k=1}^{2M-1} \alpha_k (\partial_2)^k R_\varphi(x, x_0)$$

which satisfies $\eta_W(x_0) = 1$, $\eta_W^{(k)} = 0$ for $k \in \{1, \dots, 2M-1\}$. \square

Remark 2. Note that the condition $D_V(t) > 0$ and $D_W(t) > 0$ are satisfied for all $t \in X$ in the case where the family

$$(1 \quad R_\varphi(\cdot, x_1) \quad \partial_2 R_\varphi(\cdot, x_1) \quad \dots \quad R_\varphi(\cdot, x_M) \quad \partial_2 R_\varphi(\cdot, x_M)) \quad (30)$$

is an *extended T-system* in the sense of [29].

3.3 The Cauchy-Binet formula

As noticed in [41], the Cauchy-Binet formula (also called *basic composition formula* [29, Ch. 1, example 8]) is a powerful tool to prove recovery results for the BLASSO. In this section, we describe how it yields sufficient conditions for the criteria exhibited above.

As explained in Section 2.4, we consider reconstruction frameworks which are determined by $\mathcal{H} = L^2(\mathcal{S}, P)$, $\varphi(x) = s \mapsto \psi(x, s)$, so that

$$R_\varphi(x, x') \stackrel{\text{def.}}{=} \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}} = \int_{\mathcal{S}} \psi(x, s) \psi(x', s) dP_S(s).$$

We denote by $P_S^{\otimes M}$ the product measure of P_S on S^M , and we define

$$\mathcal{T}_S^M \stackrel{\text{def.}}{=} \{(s_1, \dots, s_M) \in S^M : s_1 < \dots < s_M\}. \quad (31)$$

Since we extensively use the Cauchy-Binet formula and variants of it in the next paragraphs, let us recall its principle. Given $\theta, \xi \in L^2(\mathcal{S}, P)$, one computes the following determinant by using linearity along each column

$$\begin{aligned} & \left| \begin{array}{ccc} \int_S \theta(t_1, s) \xi(u_1, s) dP_S(s) & \dots & \int_S \theta(t_1, s) \xi(u_M, s) dP_S(s) \\ \vdots & & \vdots \\ \int_S \theta(t_M, s) \xi(u_1, s) dP_S(s) & \dots & \int_S \theta(t_M, s) \xi(u_M, s) dP_S(s) \end{array} \right| \\ &= \int_S \dots \int_S \left| \begin{array}{ccc} \theta(t_1, s_1) \xi(u_1, s_1) & \dots & \theta(t_1, s_M) \xi(u_M, s_M) \\ \vdots & & \vdots \\ \theta(t_M, s_1) \xi(u_1, s_1) & \dots & \theta(t_M, s_M) \xi(u_M, s_M) \end{array} \right| dP_S(s_1) \dots dP_S(s_M) \\ &= \int_{S^M} \xi(u_1, s_1) \dots \xi(u_M, s_M) \left| \begin{array}{ccc} \theta(t_1, s_1) & \dots & \theta(t_1, s_M) \\ \vdots & & \vdots \\ \theta(t_M, s_1) & \dots & \theta(t_M, s_M) \end{array} \right| dP_S^{\otimes M}(s_1, \dots, s_M). \end{aligned} \quad (32)$$

Now, the determinant vanishes if at least two s_i values are equal. Removing such M -plets, the remaining M -plets may be written as $(s_{\sigma(1)}, \dots, s_{\sigma(M)})$ for some vector

$(s_1, \dots, s_M) \in \mathcal{T}_S^M$ and some permutation σ of $\{1, \dots, M\}$. Splitting the domain S^M along all such permutations and reordering the columns of the determinant in θ , we obtain that (32) is equal to

$$\int_{\mathcal{T}_S^M} \left| \begin{array}{ccc} \xi(u_1, s_1) & \cdots & \xi(u_1, s_M) \\ \vdots & & \vdots \\ \xi(u_M, s_1) & \cdots & \xi(u_M, s_M) \end{array} \right| \left| \begin{array}{ccc} \theta(t_1, s_1) & \cdots & \theta(t_1, s_M) \\ \vdots & & \vdots \\ \theta(t_M, s_1) & \cdots & \theta(t_M, s_M) \end{array} \right| dP_S^{\otimes M}(s_1, \dots, s_M). \quad (33)$$

We note that the integration domain satisfies $P_S^{\otimes M}(\mathcal{T}_S^M) > 0$ if and only if $\text{supp}(P_S)$ has at least M points.

Now, we introduce several determinants which are helpful to prove that the conditions of Theorem 3 hold.

$$A_V(s_1, \dots, s_{2M}) \stackrel{\text{def.}}{=} \left| \begin{array}{ccccc} \psi(x_1, s_1) & \partial_1 \psi(x_1, s_1) & \cdots & \psi(x_M, s_1) & \partial_1 \psi(x_M, s_1) \\ \vdots & \vdots & & \vdots & \vdots \\ \psi(x_1, s_{2M}) & \partial_1 \psi(x_1, s_{2M}) & \cdots & \psi(x_M, s_{2M}) & \partial_1 \psi(x_M, s_{2M}) \end{array} \right|, \quad (34)$$

$$\text{and } B_V(t, s_1, \dots, s_{2M}) \stackrel{\text{def.}}{=} \frac{1}{\prod_{i=1}^M (t - x_i)^2} \left| \begin{array}{cccc} 1 & \psi(t, s_1) & \cdots & \psi(t, s_{2M}) \\ 1 & \psi(x_1, s_1) & \cdots & \psi(x_1, s_{2M}) \\ 0 & \partial_1 \psi(x_1, s_1) & \cdots & \partial_1 \psi(x_1, s_{2M}) \\ \vdots & \vdots & & \vdots \\ 1 & \psi(x_M, s_1) & \cdots & \psi(x_M, s_{2M}) \\ 0 & \partial_1 \psi(x_M, s_1) & \cdots & \partial_1 \psi(x_M, s_{2M}) \end{array} \right|. \quad (35)$$

Similarly to (24), B_V may be extended by continuity at each $t = x_i$ with a determinant whose first row is $(0 \quad ((\partial_1)^2 \psi(x_i, s_j))_{1 \leq j \leq 2M})$.

By an elementary variant of the Cauchy-Binet formula, one may prove that

$$\det(\Gamma_x^* \Gamma_x) = \int_{\mathcal{T}_S^{2M}} (A_V(s_1, \dots, s_{2M}))^2 dP_S^{\otimes 2M}(s_1, \dots, s_{2M}),$$

$$D_V(t) = \int_{\mathcal{T}_S^{2M}} A_V(s_1, \dots, s_{2M}) B_V(t, s_1, \dots, s_{2M}) dP_S^{\otimes 2M}(s_1, \dots, s_{2M}).$$

We obtain

Proposition 1. *If $\text{card}(\text{supp}(P_S)) \geq 2M$, and*

$$P_S^{\otimes 2M}\text{-a.e. } (s_1, \dots, s_{2M}) \in \mathcal{T}_S^{2M}, \quad A_V(s_1, \dots, s_{2M}) > 0, \quad (36)$$

then $\Gamma_{\mathbf{x}}$ has full rank.

For all $t \in X$, if moreover

$$P_S^{\otimes 2M}\text{-a.e. } (s_1, \dots, s_{2M}) \in \mathcal{T}_S^{2M}, \quad B_V(t, s_1, \dots, s_{2M}) > 0, \quad (37)$$

then $D_V(t) > 0$.

Similarly, we may define

$$A_W(s_1, \dots, s_{2M}) \stackrel{\text{def.}}{=} \begin{vmatrix} \psi(x_0, s_1) & \partial_1 \psi(x_0, s_1) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_1) \\ \vdots & \vdots & & \vdots \\ \psi(x_0, s_{2M}) & \partial_1 \psi(x_0, s_{2M}) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_{2M}) \end{vmatrix}, \quad (38)$$

$$\text{and } B_W(t, s_1, \dots, s_{2M}) \stackrel{\text{def.}}{=} \frac{1}{(t - x_0)^{2M}} \begin{vmatrix} 1 & \psi(t, s_1) & \cdots & \psi(t, s_{2M}) \\ 1 & \psi(x_0, s_1) & \cdots & \psi(x_0, s_{2M}) \\ 0 & \partial_1 \psi(x_0, s_1) & \cdots & \partial_1 \psi(x_0, s_{2M}) \\ \vdots & \vdots & & \vdots \\ 0 & (\partial_1)^{2M-1} \psi(x_0, s_1) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_{2M}) \end{vmatrix}. \quad (39)$$

As above, one may prove that

$$\det(\Psi_{2M-1}^* \Psi_{2M-1}) = \int_{\mathcal{T}_S^{2M}} (A_W(s_1, \dots, s_{2M}))^2 dP_S^{\otimes 2M}(s_1, \dots, s_{2M}),$$

$$D_W(t) = \int_{\mathcal{T}_S^{2M}} A_W(s_1, \dots, s_{2M}) B_W(t, s_1, \dots, s_{2M}) dP_S^{\otimes 2M}(s_1, \dots, s_{2M}).$$

thus

Proposition 2. *If $\text{card}(\text{supp}(P_S)) \geq 2M$, and*

$$P_S^{\otimes 2M}\text{-a.e. } (s_1, \dots, s_{2M}) \in \mathcal{T}_S^{2M}, \quad A_W(s_1, \dots, s_{2M}) > 0, \quad (40)$$

then $\Gamma_{\mathbf{x}}$ has full rank.

For all $t \in X$, if moreover

$$P_S^{\otimes 2M}\text{-a.e. } (s_1, \dots, s_{2M}) \in \mathcal{T}_S^{2M}, \quad B_W(t, s_1, \dots, s_{2M}) > 0, \quad (41)$$

then $D_W(t) > 0$.

3.4 Extension to the unconstrained Lasso

While the main focus of this paper is the Lasso with positivity constraint, let us note that the characterization of the NDSC provided in Theorem 3 can be adapted to the unconstrained Lasso. If $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, let $I^+ = \{i \in \{1, \dots, M\} : a_i > 0\}$ and $I^- = \{i \in \{1, \dots, M\} : a_i < 0\}$. The Non-Degenerate Source Condition on η_V is that

$$\forall i \in I^+, \quad \eta_V(x_i) = 1, \quad \eta_V''(x_i) < 0, \quad (42)$$

$$\forall i \in I^-, \quad \eta_V(x_i) = -1, \quad \eta_V''(x_i) > 0, \quad (43)$$

$$\forall t \in X \setminus \{x_i\}_{i=1}^M, \quad |\eta_V(t)| < 1. \quad (44)$$

Using the same arguments as above, it is possible to prove that the Non-Degenerate Source Condition is equivalent to

$$\forall t \in X, \quad D_V^+(t) > 0 \quad \text{and} \quad D_V^-(t) < 0. \quad (45)$$

where

$$\begin{aligned}
D_V^+(t) &\stackrel{\text{def.}}{=} \frac{2}{\prod_{i \in I^+} (t - x_i)^2} \\
&\times \begin{vmatrix} 1 & R_\varphi(t, x_1) & \partial_2 R_\varphi(t, x_1) & \cdots & R_\varphi(t, x_M) & \partial_2 R_\varphi(t, x_M) \\ \text{sign}(a_1) & R_\varphi(x_1, x_1) & \partial_2 R_\varphi(x_1, x_1) & \cdots & R_\varphi(x_1, x_M) & \partial_2 R_\varphi(x_1, x_M) \\ 0 & \partial_1 R_\varphi(x_1, x_1) & \partial_1 \partial_2 R_\varphi(x_1, x_1) & \cdots & \partial_1 R_\varphi(x_1, x_M) & \partial_1 \partial_2 R_\varphi(x_1, x_M) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \text{sign}(a_M) & R_\varphi(x_M, x_1) & \partial_2 R_\varphi(x_M, x_1) & \cdots & R_\varphi(x_M, x_M) & \partial_2 R_\varphi(x_M, x_M) \\ 0 & \partial_1 R_\varphi(x_M, x_1) & \partial_1 \partial_2 R_\varphi(x_M, x_1) & \cdots & \partial_1 R_\varphi(x_M, x_M) & \partial_1 \partial_2 R_\varphi(x_M, x_M) \end{vmatrix}, \\
D_V^-(t) &\stackrel{\text{def.}}{=} \frac{2}{\prod_{i \in I^-} (t - x_i)^2} \\
&\times \begin{vmatrix} -1 & R_\varphi(t, x_1) & \partial_2 R_\varphi(t, x_1) & \cdots & R_\varphi(t, x_M) & \partial_2 R_\varphi(t, x_M) \\ \text{sign}(a_1) & R_\varphi(x_1, x_1) & \partial_2 R_\varphi(x_1, x_1) & \cdots & R_\varphi(x_1, x_M) & \partial_2 R_\varphi(x_1, x_M) \\ 0 & \partial_1 R_\varphi(x_1, x_1) & \partial_1 \partial_2 R_\varphi(x_1, x_1) & \cdots & \partial_1 R_\varphi(x_1, x_M) & \partial_1 \partial_2 R_\varphi(x_1, x_M) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \text{sign}(a_M) & R_\varphi(x_M, x_1) & \partial_2 R_\varphi(x_M, x_1) & \cdots & R_\varphi(x_M, x_M) & \partial_2 R_\varphi(x_M, x_M) \\ 0 & \partial_1 R_\varphi(x_M, x_1) & \partial_1 \partial_2 R_\varphi(x_M, x_1) & \cdots & \partial_1 R_\varphi(x_M, x_M) & \partial_1 \partial_2 R_\varphi(x_M, x_M) \end{vmatrix}.
\end{aligned}$$

4 The case of Laplace and Gaussian measurements

Now, we discuss the applicability of Theorem 3 to Laplace and Gaussian measurements.

4.1 Laplace measurements $\psi(x, s) = e^{-xs}$

It is standard that the exponential kernel $K(x, y) = e^{xy}$ is *extended totally positive* (see [28, Ch. 3]), that is for all $N \in \mathbb{N}^*$, all real numbers $x_1 < \dots < x_N$, $x'_1 < \dots < x'_N$, there holds $\det((K(x_i, x'_j))_{1 \leq i, j \leq N}) > 0$. If $x_i = x_{i+1} = \dots = x_{i+r-1}$, then one may replace the corresponding rows with the successive derivatives

$$\begin{pmatrix} K(x_i, x'_1) & \cdots & K(x_i, x'_N) \\ \partial_1 K(x_i, x'_1) & \cdots & \partial_1 K(x_i, x'_N) \\ \vdots & & \vdots \\ (\partial_1)^{r-1} K(x_i, x'_1) & \cdots & (\partial_1)^{r-1} K(x_i, x'_N) \end{pmatrix} \quad (46)$$

and still get a strictly positive quantity. The same holds w.r.t x' .

Changing y into $-s$, setting $\psi(t, s) \stackrel{\text{def.}}{=} K(t, -s)$ and reordering the rows, we see that the sign of $\det((\exp(-t_i s_j))_{1 \leq i, j \leq N})$ is $(-1)^{\frac{N(N-1)}{2}}$ if $t_1 \leq \dots \leq t_N$ and $s_1 \leq \dots \leq s_N$ (and the usual adaptation if some t_i or s_j values are equal).

Our first result in this direction is the following.

Corollary 1. *Let $X = [c, +\infty)$ with $c \geq 0$, $S \subset (0, +\infty)$ and P_S be a positive measure such that $\int_S (1 + |s|)^{4M} e^{-2cs} dP_S(s) < +\infty$.*

If $\psi(x, s) = e^{-xs}$ and $\text{card}(\text{supp}(P_S)) \geq 2M$, the following holds.

- *If $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \text{int}(X)$ and $a_i > 0$ for all i , then m_0 satisfies the Non-Degenerate Source Condition.*
- *If $x_0 \in \text{int}(X)$, then the point x_0 is $(2M - 1)$ Non Degenerate.*

The proof is given in Appendix A.1.

Remark 3. One might wonder why we impose $S \subset (0, +\infty)$ rather than $S \subset [0, +\infty)$. It is due to the artificial point s_0 which is introduced. It is possible to choose $S = [0, +\infty)$ and $0 \in \text{supp}(P_S)$, but then $2M$ measurements are not sufficient anymore, since for $s_1 = 0 = s_0$, the function $t \mapsto B_V(t, s_1, \dots, s_{2M})$ vanishes identically. As a result, η_V (resp. η_W) is identically equal to 1 and the (NDSC) does not hold.

In that case, the conclusion of Proposition 1 still holds with the stronger condition that $\text{card}(\text{supp}(P_S)) \geq 2M + 1$; everything happens as if the measurement at $s = 0$ were ignored.

4.2 Gaussian kernel $\psi(x, s) = e^{-(x-s)^2}$

Now, we deal with the more involved case of the Gaussian kernel. As noted in [41], a family of the form $\{1, e^{-(\cdot-x_1)^2}, \dots, e^{-(\cdot-x_M)^2}\}$ does not form a T -system, hence the sufficient condition edicted in [11] for identifiability does not hold. For that reason, the authors of [41] were led to introduce a weighted total variation (equivalent to some renormalization of the impulse responses) to provide identifiability.

In our context, $B_V(t, s_1, \dots, s_{2M}) > 0$ does not always hold, hence it is not possible to assert that the conditions of Theorem 3 hold regardless of the sampling (S, P_S) . In fact, the experiments shown in Figures 2 and 3 indicate that for some samplings measures P_S , the problem is not support stable (is not clear to us whether or not m_0 is identifiable in such cases).

However, as described below, it is possible to assert support stability in two cases which are worthy of interest. The first one is the regular convolution with a Gaussian kernel, without sampling, or with some sampling which approximates the Lebesgue measure (in a suitable weak sense). The second one is when the sampling is confined into a small interval.

4.2.1 Full sampling, or almost

When fully observing the convolution of m_0 with some Gaussian kernel, *i.e.* when

$$\varphi(x) : \mathbb{R} \rightarrow L^2(\mathbb{R}), \quad s \mapsto \int_{\mathbb{R}} e^{-(x-s)^2} dm_0(x), \quad (47)$$

the operator Φ is injective, hence it is clear that m_0 is the unique solution to $(\mathcal{P}_0(\Phi m_0))$. One might wonder, however, whether this recovery is support stable.

Proposition 3 (Fully sampled Gaussian kernel). *Let P_S be the Lebesgue measure on \mathbb{R} , $X = \mathbb{R}$ and $\psi(x, s) \stackrel{\text{def.}}{=} e^{-(x-s)^2}$. Then*

- Any measure $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \mathbb{R}$ and $a_i > 0$ for all i , satisfies the Non-Degenerate Source Condition.
- For any point $x_0 \in \mathbb{R}$, then the point x_0 is $(2M - 1)$ Non Degenerate.

A proof of Proposition 3 is given in Appendix A.2. Let us note that this proposition is a consequence of the following technical lemma proved in [41].

Lemma 4 ([41, Lemma 2.7]). *Let $d_1, \dots, d_r \in \mathbb{N}$ and $s_1 < \dots < s_r$ be real numbers. The function*

$$t \mapsto \sum_{i=1}^r (a_{i,0} + a_{i,1}t + \dots + a_{i,(2d_i-1)}t^{(2d_i-1)})e^{s_i t} + e^{t^2} \quad (48)$$

has at most $2(d_1 + \dots + d_r)$ zeros, counting multiplicities.

As a consequence of Proposition 3, we also obtain non-degeneracy results for measures with finite total mass which are sufficiently close to the uniform Lebesgue measure. More precisely, let us assume that

$$\lim_{n \rightarrow +\infty} \max_{0 \leq k, \ell \leq 2M} \sup_{x \in X} \left| \int_{\mathbb{R}} x^k s^\ell e^{-(x-s)^2 - (x_i-s)^2} dP_{S,n}(s) - \int_{\mathbb{R}} x^k s^\ell e^{-(x-s)^2 - (x_i-s)^2} ds \right| = 0. \quad (49)$$

Proposition 4. *Let $X \subseteq \mathbb{R}$, $(P_{S,n})_{n \in \mathbb{N}}$ be a sequence of positive measures with finite total mass such that (49) holds. Let $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \text{int}(X)$ and $a_i > 0$ for all i (resp. let $x_0 \in \text{int}(X)$).*

Then, provided n is large enough,

- i) m_0 satisfies the Non-Degenerate Source Condition,*
- ii) The point x_0 is $2M - 1$ Non-Degenerate,*

for the reconstruction framework defined by $P_{S,n}$ and $\psi(x, s) = e^{-(x-s)^2}$.

A proof of Proposition 4 is given in Appendix A.3.

4.2.2 Confined sampling

Now, we return to the framework of Section 3.3 with some sampling measure P_S . We show that if the support of P_S is confined in some sufficiently small interval, stable recovery of the support is possible regardless of the spikes separation.

Proposition 5. *Let $X \subseteq \mathbb{R}$ be an interval, $s_* \in \mathbb{R}$ and P_S be a positive measure such that $\text{card}(\text{supp}(P_S)) \geq 2M$. There exists $r > 0$ such that if $\text{supp} P_S \subset [s_* - r, s_* + r]$, then for any measure $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \text{int}(X)$ and $a_i > 0$ for all i (resp. for any $x_0 \in \text{int}(X)$),*

- i) m_0 satisfies the Non-Degenerate Source Condition,*
- ii) The point x_0 is $2M - 1$ Non-Degenerate,*

in the reconstruction framework defined by P_S and $\psi(x, s) = e^{-(x-s)^2}$.

A proof of Proposition 5 is given in Appendix A.4.

5 Renormalizing the atoms

In this last section, we discuss the situation when renormalizing the atoms, which amounts to the setting of [41]. Instead of directly using φ and the corresponding correlation $R_\varphi(x, y)$ in $(\mathcal{P}_0(\Phi m_0))$ and $(\mathcal{P}_\lambda(y))$, one may wish to renormalize the atoms and consider $\bar{\varphi}(x) \stackrel{\text{def.}}{=} \frac{1}{n(x)} \varphi(x)$ and the corresponding correlation $R_{\bar{\varphi}}(x, y) = \frac{1}{n(x)n(y)} R_\varphi(x, y)$, where $x \mapsto n(x)$ is a smooth (strictly) positive function. For instance $n(x) = \sqrt{R_\varphi(x, x)}$ yields atoms $\bar{\varphi}(x)$ with the same norm in \mathcal{H} .

It is then important to understand when the Non-Degenerate Source Condition holds for $R_{\bar{\varphi}}$.

For notational simplicity, we define

$$(\bar{u}_0, \bar{u}_1, \bar{u}_2, \dots, \bar{u}_{2M-1}, \bar{u}_{2M}) \stackrel{\text{def.}}{=} (1, R_{\bar{\varphi}}(\cdot, x_1), \partial_2 R_{\bar{\varphi}}(\cdot, x_1), \dots, R_{\bar{\varphi}}(\cdot, x_M), \partial_2 R_{\bar{\varphi}}(\cdot, x_M)). \quad (50)$$

and we consider the associated \bar{D}_V defined through (22).

We notice that

$$\partial_2 R_{\bar{\varphi}}(t, x_j) = \frac{1}{n(t)n(x_j)} \partial_2 R_{\varphi}(t, x_j) + \frac{R_{\varphi}(t, x_j)}{n(t)} \frac{d}{dx_j} \left(\frac{1}{n(x_j)} \right) \quad (51)$$

therefore by subtracting a multiple of each even column to its successor, we see that for all $t \in X \setminus \{x_i\}_{i=1}^M$,

$$\begin{vmatrix} 1 & R_{\bar{\varphi}}(t, x_1) & \partial_2 R_{\bar{\varphi}}(t, x_1) & \dots & R_{\bar{\varphi}}(t, x_M) & \partial_2 R_{\bar{\varphi}}(t, x_M) \\ 1 & R_{\bar{\varphi}}(x_1, x_1) & \partial_2 R_{\bar{\varphi}}(x_1, x_1) & \dots & R_{\bar{\varphi}}(x_1, x_M) & \partial_2 R_{\bar{\varphi}}(x_1, x_M) \\ 0 & \partial_1 R_{\bar{\varphi}}(x_1, x_1) & \partial_1 \partial_2 R_{\bar{\varphi}}(x_1, x_1) & \dots & \partial_1 R_{\bar{\varphi}}(x_1, x_M) & \partial_1 \partial_2 R_{\bar{\varphi}}(x_1, x_M) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & R_{\bar{\varphi}}(x_M, x_1) & \partial_2 R_{\bar{\varphi}}(x_M, x_1) & \dots & R_{\bar{\varphi}}(x_M, x_M) & \partial_2 R_{\bar{\varphi}}(x_M, x_M) \\ 0 & \partial_1 R_{\bar{\varphi}}(x_M, x_1) & \partial_1 \partial_2 R_{\bar{\varphi}}(x_M, x_1) & \dots & \partial_1 R_{\bar{\varphi}}(x_M, x_M) & \partial_1 \partial_2 R_{\bar{\varphi}}(x_M, x_M) \end{vmatrix} \\ = \frac{1}{\left(\prod_{j=1}^M n(x_j) \right)^2} \begin{vmatrix} 1 & \frac{R_{\varphi}(t, x_1)}{n(t)} & \frac{\partial_2 R_{\varphi}(t, x_1)}{n(t)} & \dots & \frac{R_{\varphi}(t, x_M)}{n(t)} & \frac{\partial_2 R_{\varphi}(t, x_M)}{n(t)} \\ 1 & \frac{R_{\varphi}(x_1, x_1)}{n(x_1)} & \frac{\partial_2 R_{\varphi}(x_1, x_1)}{n(x_1)} & \dots & \frac{R_{\varphi}(x_1, x_M)}{n(x_1)} & \frac{\partial_2 R_{\varphi}(x_1, x_M)}{n(x_1)} \\ 0 & \frac{\partial_1 R_{\varphi}(x_1, x_1)}{n(x_1)} & \frac{\partial_1 \partial_2 R_{\varphi}(x_1, x_1)}{n(x_1)} & \dots & \frac{\partial_1 R_{\varphi}(x_1, x_M)}{n(x_1)} & \frac{\partial_1 \partial_2 R_{\varphi}(x_1, x_M)}{n(x_1)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & \frac{R_{\varphi}(x_M, x_1)}{n(x_M)} & \frac{\partial_2 R_{\varphi}(x_M, x_1)}{n(x_M)} & \dots & \frac{R_{\varphi}(x_M, x_M)}{n(x_M)} & \frac{\partial_2 R_{\varphi}(x_M, x_M)}{n(x_M)} \\ 0 & \frac{\partial_1 R_{\varphi}(x_M, x_1)}{n(x_M)} & \frac{\partial_1 \partial_2 R_{\varphi}(x_M, x_1)}{n(x_M)} & \dots & \frac{\partial_1 R_{\varphi}(x_M, x_M)}{n(x_M)} & \frac{\partial_1 \partial_2 R_{\varphi}(x_M, x_M)}{n(x_M)} \end{vmatrix} \quad (52)$$

$$= \frac{1}{n(t) \left(\prod_{j=1}^M n(x_j) \right)^4} \quad (53)$$

$$\times \begin{vmatrix} n(t) & R_{\varphi}(t, x_1) & \partial_2 R_{\varphi}(t, x_1) & \dots & R_{\varphi}(t, x_M) & \partial_2 R_{\varphi}(t, x_M) \\ n(x_1) & R_{\varphi}(x_1, x_1) & \partial_2 R_{\varphi}(x_1, x_1) & \dots & R_{\varphi}(x_1, x_M) & \partial_2 R_{\varphi}(x_1, x_M) \\ 0 & \partial_1 R_{\varphi}(x_1, x_1) & \partial_1 \partial_2 R_{\varphi}(x_1, x_1) & \dots & \partial_1 R_{\varphi}(x_1, x_M) & \partial_1 \partial_2 R_{\varphi}(x_1, x_M) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ n(x_M) & R_{\varphi}(x_M, x_1) & \partial_2 R_{\varphi}(x_M, x_1) & \dots & R_{\varphi}(x_M, x_M) & \partial_2 R_{\varphi}(x_M, x_M) \\ 0 & \partial_1 R_{\varphi}(x_M, x_1) & \partial_1 \partial_2 R_{\varphi}(x_M, x_1) & \dots & \partial_1 R_{\varphi}(x_M, x_M) & \partial_1 \partial_2 R_{\varphi}(x_M, x_M) \end{vmatrix} \quad (54)$$

The above matrix, divided by $\prod_{i=1}^M (t - x_i)^2$ is a infinitesimal version of the matrix $\Lambda(t, x_1, x_1, \dots, x_M, x_M)$ which appears in the DETERMINANTAL condition of [41]. We note that it is not specific to the L^1 norm, as n could be any (positive smooth) normalizing factor.

Remark 4. If we define $\bar{\Gamma}_{\mathbf{x}} \stackrel{\text{def.}}{=} (\bar{\varphi}(x_1) \dots \bar{\varphi}(x_M) \bar{\varphi}'(x_1) \dots \bar{\varphi}'(x_M))$, we note in view of (51) that $\det(\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})$ and $\det(\bar{\Gamma}_{\mathbf{x}}^* \bar{\Gamma}_{\mathbf{x}})$ are equal up to a (strictly) positive multiplicative factor. Hence, when working with normalized atoms, it is sufficient to check that $\Gamma_{\mathbf{x}}$ has full rank (on the unnormalized family). Similarly, if we define $\bar{\Psi}_k \stackrel{\text{def.}}{=} (\bar{\varphi}(x_0) \bar{\varphi}'(x_0) \dots \bar{\varphi}^{(k)}(x_0))$, we note that $\det(\Psi_k^* \Psi_k)$ and $\det(\bar{\Psi}_k^* \bar{\Psi}_k)$ are equal up to a (strictly) positive multiplicative factor. Hence, when working with normalized atoms, it is sufficient to check that Ψ_k has full rank (on the unnormalized family).

5.1 Cauchy-Binet formula

In the rest of this section, to simplify things, we assume that P_S is a finite measure, *i.e.* $P_S(S) < +\infty$. This ensures that the constant function $s \mapsto 1$ is in \mathcal{H} (and $\psi(t, \cdot) \in L^1(\mathcal{S}, P)$ for all $t \in X$).

To ensure that (54) is positive, one may use the same decomposition as in [41] that it is equal up to a positive factor to

$$\int_{\mathcal{T}_S^{2M+1}} \bar{A}_V(s_0, \dots, s_{2M}) \bar{B}_V(t, s_0, \dots, s_{2M}) dP_S^{\otimes 2M+1}(s_0, \dots, s_{2M}), \quad (55)$$

where

$$\bar{A}_V(s_0, \dots, s_{2M}) \stackrel{\text{def.}}{=} \begin{vmatrix} 1 & \psi(x_1, s_0) & \partial_1 \psi(x_1, s_0) & \cdots & \psi(x_M, s_0) & \partial_1 \psi(x_M, s_0) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & \psi(x_1, s_{2M}) & \partial_1 \psi(x_1, s_{2M}) & \cdots & \psi(x_M, s_{2M}) & \partial_1 \psi(x_M, s_{2M}) \end{vmatrix} \quad (56)$$

$$\bar{B}_V(t, s_0, \dots, s_{2M}) \stackrel{\text{def.}}{=} \frac{2}{\prod_{i=1}^M (t - x_i)^2} \begin{vmatrix} \psi(t, s_0) & \cdots & \psi(t, s_{2M}) \\ \psi(x_1, s_0) & \cdots & \psi(x_1, s_{2M}) \\ \partial_1 \psi(x_1, s_{2M}) & \cdots & \partial_1 \psi(x_1, s_{2M}) \\ \vdots & & \vdots \\ \psi(x_M, s_0) & \cdots & \psi(x_M, s_{2M}) \\ \partial_1 \psi(x_M, s_{2M}) & \cdots & \partial_1 \psi(x_M, s_{2M}) \end{vmatrix} \quad (57)$$

We obtain

Proposition 6. Assume that $P_S(S) < +\infty$ and let $n(t) \stackrel{\text{def.}}{=} \langle 1, \varphi(t) \rangle = \int_S \psi(t, s) dP_S(s)$. Assume moreover that $n(t) > 0$ for all $t \in X$.

1. If $\text{card}(\text{supp}(P_S)) = 2M$, then (54) is identically zero. If moreover $\Gamma_{\mathbf{x}}$ has full rank, η_V is identically 1.
2. If $\text{card}(\text{supp}(P_S)) \geq 2M+1$, and $\bar{A}_V(s_0, \dots, s_{2M}) > 0$ for $P_S^{\otimes 2M+1}$ -a.e. $(s_0, \dots, s_{2M}) \in \mathcal{T}_S^{2M+1}$, then $\bar{\Gamma}_{\mathbf{x}}$ has full column rank.
If, additionally, for $t \in X$, $P_S^{\otimes 2M+1}$ -a.e. $(s_0, \dots, s_{2M}) \in \mathcal{T}_S^{2M+1}$, $\bar{B}_V(t, s_0, \dots, s_{2M}) > 0$, then $\bar{D}_V(t) > 0$.

We skip the proof as it follows directly from (58). It should be noted that the L^1 -normalization of the atoms “spoils” one measurement: one needs $2M+1$ instead of $2M$ to get identifiability or stability.

Again, there is a variant for the limit $(x_1, \dots, x_M) \rightarrow (x_0, \dots, x_0)$. It relies on the decomposition

$$\int_{\mathcal{T}_S^{2M+1}} \bar{A}_W(s_0, \dots, s_{2M}) \bar{B}_W(t, s_0, \dots, s_{2M}) dP_S^{\otimes 2M+1}(s_0, \dots, s_{2M}), \quad (58)$$

where

$$\bar{A}_W(s_0, \dots, s_{2M}) \stackrel{\text{def.}}{=} \begin{vmatrix} 1 & \psi(x_0, s_0) & \partial_1 \psi(x_0, s_0) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_0) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \psi(x_0, s_{2M}) & \partial_1 \psi(x_0, s_{2M}) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_{2M}) \end{vmatrix} \quad (59)$$

$$\bar{B}_W(t, s_0, \dots, s_{2M}) \stackrel{\text{def.}}{=} \frac{(2M)!}{\prod_{i=1}^M (t - x_i)^2} \begin{vmatrix} \psi(t, s_0) & \cdots & \psi(t, s_{2M}) \\ \psi(x_0, s_0) & \cdots & \psi(x_0, s_{2M}) \\ \partial_1 \psi(x_0, s_{2M}) & \cdots & \partial_1 \psi(x_0, s_{2M}) \\ \vdots & & \vdots \\ (\partial_1)^{2M-1} \psi(x_0, s_{2M}) & \cdots & (\partial_1)^{2M-1} \psi(x_0, s_{2M}) \end{vmatrix} \quad (60)$$

The conclusions of Proposition 6 hold for η_W with the obvious adaptations.

5.2 L^1 -normalized Laplace measurements

As a consequence we obtain the following result for L^1 -normalized Laplace measurements, *i.e.* $\bar{\varphi}(x) : s \mapsto \frac{1}{n(x)}\psi(x, s)$ where $\psi(x, s) = e^{-xs}$.

Corollary 2. *Let $X = [c, +\infty)$ where $c \geq 0$ and let P_S be a positive measure on $S \subseteq [0, +\infty)$ such that $\int_S (1 + |s|)^{4M} e^{-2cs} dP_S(s) < +\infty$ with $\text{card}(\text{supp}(P_S)) \geq 2M + 1$,*

- *If $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \text{int}(X)$ and $a_i > 0$ for all i , then m_0 satisfies the Non-Degenerate Source Condition,*
- *If $x_0 \in \text{int}(X)$, then the point x_0 is $(2M - 1)$ Non Degenerate,*

for the reconstruction framework defined by P_S and $\bar{\psi}(x, s) = \frac{1}{n(x)}\psi(x, s)$, where $\psi(x, s) = e^{-xs}$.

The proof is straightforward, introducing the point $x = 0$ and using the extended total positivity of the exponential kernel.

5.3 Arbitrary sampling for L^1 normalized Gaussian measurements

The case of L^1 normalized Gaussian measurements, *i.e.* $\bar{\varphi}(x) : s \mapsto \frac{1}{n(x)}\psi(x, s)$, where $\psi(x, s) = e^{-(x-s)^2}$ for all $(x, s) \in \mathbb{R}^2$, and $n(x) = \int_S \psi(x, s) dP_S(s)$, is studied in [41]. The authors show that the alternative $(\eta_V < 1 \text{ on } X \setminus \{x_i\}_{i=1}^M)$ or $(\eta_V > 1 \text{ on } X \setminus \{x_i\}_{i=1}^M)$ holds, and they deduce the identifiability of non-negative sums of M Dirac masses. However it is not clear which of $\eta_V > 1$ or $\eta_V < 1$ holds, and thus, whether the support is stable at low noise or not.

In fact, their proof contains all the necessary ingredients, as they show that both \bar{A}_V and \bar{B}_V are nonzero, and similarly for \bar{A}_W and \bar{B}_W . By the same arguments as in the proof of Proposition 3, one may prove that they are in fact positive and obtain

Corollary 3. *Let P_S be a finite positive measure on $S \subseteq \mathbb{R}$, with $\text{card}(\text{supp}(P_S)) \geq 2M + 1$.*

- *If $m_0 = \sum_{i=1}^M a_i \delta_{x_i}$, with $x_i \in \text{int}(X)$ and $a_i > 0$ for all i , then m_0 satisfies the Non-Degenerate Source Condition*
- *If $x_0 \in \text{int}(X)$, then the point x_0 is $(2M - 1)$ Non Degenerate*

for the reconstruction framework defined by $\bar{\psi}(x, s) = \frac{1}{n(x)}\psi(x, s)$ with $\psi(x, s) = e^{-(x-s)^2}$ for all $(x, s) \in \mathbb{R}^2$.

Conclusion

In this work, we have provided a necessary and sufficient condition for the Non-Degenerate Source Condition which ensures support stability for the total variation sparse recovery of measures. We have proved that this condition holds for the partial Laplace transform, regardless of the repartition of the samples. In the case of the convolution with a Gaussian filter, we have shown that the proposed conditions hold in the case of the full observation of the convolution, as well as for samplings which approximate the Lebesgue measure. Additionally such conditions hold in the case of *confined samplings*, when the convolution is observed in some small interval.

The main purpose of the present work is to lay the theoretical foundations of super-resolution with Laplace observations [14] in view of the forthcoming paper [14] which deals with Total Internal Reflection Microscopy (TIRF).

Acknowledgements

The author would like to thank Gabriel Peyré, Quentin Denoyelle and Emmanuel Soubies for stimulating discussions about this work, in particular for suggesting the development around the unsampled Gaussian convolution.

A Proofs of Section 4

A.1 Proof of Corollary 1

Proof. We apply Proposition 1 (resp. Proposition 2) and Theorem 3, noting that the conclusion of Proposition 1 still holds if we only assume that the product of the two determinants $A_V(s_1, \dots, s_{2M})$ and $B_V(t, s_1, \dots, s_M)$ is (strictly) positive.

The condition $\int_S (1 + |s|)^{4M} e^{-2cs} dP_S(s) < +\infty$ ensures that (21) holds for $0 \leq k \leq 2M$. As noted above, the exponential kernel is *extended totally positive*, if $0 < x_1 < \dots < x_M$ (resp. $0 < x_0$), and $0 < s_1 < \dots < s_{2M}$, we get

$$\begin{aligned} 0 &< (-1)^{M(2M-1)} \begin{vmatrix} \psi(x_1, s_1) & \dots & \partial_1 \psi(x_M, s_1) \\ \vdots & & \vdots \\ \psi(x_1, s_{2M}) & \dots & \partial_1 \psi(x_M, s_{2M}) \end{vmatrix} \\ &= (-1)^{M(2M-1)} A_V(s_1, \dots, s_{2M}), \\ (\text{resp.}) \quad 0 &< (-1)^{M(2M-1)} \begin{vmatrix} \psi(x_0, s_1) & \dots & (\partial_1)^{2M-1} \psi(x_0, s_1) \\ \vdots & & \vdots \\ \psi(x_0, s_{2M}) & \dots & (\partial_1)^{2M-1} \psi(x_0, s_{2M}) \end{vmatrix} \\ &= (-1)^{M(2M-1)} A_W(s_1, \dots, s_{2M}). \end{aligned}$$

As for the other determinant, introducing $s_0 = 0$, we obtain

$$\begin{aligned} 0 &< \frac{(-1)^{M(2M+1)}}{\prod_{i=1}^M (t - x_i)^2} \begin{vmatrix} \psi(t, s_0) & \psi(t, s_1) & \dots & \psi(t, s_{2M}) \\ \psi(x_1, s_0) & \partial \psi(x_1, s_1) & \dots & \psi(x_1, s_{2M}) \\ \partial_1 \psi(x_1, s_0) & \partial_1 \partial \psi(x_1, s_1) & \dots & \partial_1 \psi(x_1, s_{2M}) \\ \vdots & \vdots & & \vdots \\ \psi(x_M, s_0) & \partial \psi(x_M, s_1) & \dots & \psi(x_M, s_{2M}) \\ \partial_1 \psi(x_M, s_0) & \partial_1 \partial \psi(x_M, s_1) & \dots & \partial_1 \psi(x_M, s_{2M}) \end{vmatrix} \\ &= (-1)^{M(2M+1)} B_V(t, s_1, \dots, s_M) \end{aligned}$$

The same holds for $t = x_i$, with the usual extension. \square

A.2 Proof of Proposition 3

Proof of Proposition 3. The correlation function is given by $R_\varphi(x, x') = e^{\frac{1}{2}(x-x')^2} R_\varphi(0, 0)$, and $\partial_2 R_\varphi(x, x') = (x' - x) e^{\frac{1}{2}(x-x')^2} R_\varphi(0, 0)$.

Setting $K(z, z') = e^{\frac{1}{2}(z-z')^2}$ we see that, up to a positive factor, $\det(\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})$ is exactly the determinant $\det(K(z_i, z'_j))$ for $z_{2i-1} = z_{2i} = x_i$ (resp. $z'_{2i-1} = z'_{2i} = x_i$) with the replacement introduced described in (46) in z and z' . Hence, by the *extended total positivity* of the Gaussian kernel (see [28]), $\det(\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}}) > 0$ and $\Gamma_{\mathbf{x}}$ has full rank.

Now, dropping the constant factors $R_\varphi(0, 0)$, let us write

$$\begin{aligned} u_0(t) &:= 1, & u_1(t) &:= e^{-(t-x_1)^2/2}, & \dots & & u_{2M-1}(t) &:= e^{-(t-x_M)^2/2}, \\ u_2(t) &:= (t - x_1) e^{-(t-x_1)^2/2}, & \dots & & u_{2M}(t) &:= (t - x_M) e^{-(t-x_M)^2/2}. \end{aligned}$$

By the Leibniz formula, we note that any function of the form

$$t \mapsto \sum_{i=0}^{2M} b_i u_i(t)$$

has the same roots (with the same multiplicities) as the function

$$t \mapsto \sum_{i=1}^M \left[(b_{2i-1} + b_{2i}(t - x_i)) e^{\frac{1}{2}x_i^2} \right] e^{x_i t} + b_0 e^{\frac{1}{2}t^2}, \quad (61)$$

If $b_0 = 0$, it is known that (61) has at most $2M - 1$ zeros (as a consequence of the extended total positivity of the Gaussian, see also [38, Ex. V.75]). If $b_0 \neq 0$, it follows from Lemma 4 that (61) has at most $2M$ zeros. As a result, the function

$$D_V : t \mapsto \frac{2}{\prod_{i=1}^M (t - x_i)^2} \begin{vmatrix} u_0(t) & u_1(t) & \cdots & u_{2M}(t) \\ u_0(x_1) & u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_0(x_1) & u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & \vdots & & \vdots \\ u_0(x_M) & u_1(x_M) & \cdots & u_{2M}(x_M) \\ u'_0(x_M) & u'_1(x_M) & \cdots & u'_{2M}(x_M) \end{vmatrix}, \quad (62)$$

which is continuous on X , does not vanish. To evaluate its sign, we let $t \rightarrow +\infty$, noting that $u_i(t) \rightarrow 0$ for $1 \leq i \leq 2M$, and we expand along the first row

$$D_V(t) = \frac{2}{\prod_{i=1}^M (t - x_i)^2} \left(\begin{vmatrix} u_1(x_1) & \cdots & u_{2M}(x_1) \\ u'_1(x_1) & \cdots & u'_{2M}(x_1) \\ \vdots & & \vdots \\ u_1(x_M) & \cdots & u_{2M}(x_M) \\ u'_1(x_M) & \cdots & u'_{2M}(x_M) \end{vmatrix} + o(1) \right) \\ > 0 \text{ for } t \text{ large enough,}$$

by the extended total positivity of the Gaussian. As a result, D_V is strictly positive on \mathbb{R} , and the (NDSC) holds.

The second conclusion is obtained by applying similar arguments to $\det(\Psi_{2M-1}^* \Psi_{2M-1})$ and to the functions

$$u_0(t) := 1, \quad u_1(t) := e^{-(t-x_0)^2/2}, \quad u_2(t) = \hat{H}_2(t) e^{-(t-x_0)^2/2}, \quad \dots \quad \hat{H}_{2M-1}(t) e^{-(t-x_0)^2/2}, \quad (63)$$

where \hat{H}_k is defined by $\hat{H}_k(u) = e^{u^2} \frac{d^k}{du^k} (e^{-u^2})$, that is (up to a $(-1)^k$ factor) the k -th Hermite polynomial (note that each \hat{H}_k has degree k). We skip the detail for brevity. \square

A.3 Proof of Proposition 4

Proof. The autocorrelation functions corresponding to the Lebesgue measure and $P_{S,n}$ are defined by

$$R_\varphi(x, x') = \int_{\mathbb{R}} e^{-(x-s)^2} e^{-(x'-s)^2} ds, \\ R_\varphi^n(x, x') = \int_{\mathbb{R}} e^{-(x-s)^2} e^{-(x'-s)^2} dP_{S,n}(s),$$

so that, denoting by \hat{H}_k the k -th Hermite polynomial as above,

$$\begin{aligned}(\partial_1)^k(\partial_2)^\ell R_\varphi(x, x') &= \int_{\mathbb{R}} \hat{H}_k(x-s) \hat{H}_\ell(x'-s) e^{-(x-s)^2} e^{-(x'-s)^2} ds, \\(\partial_1)^k(\partial_2)^\ell R_\varphi^n(x, x') &= \int_{\mathbb{R}} \hat{H}_k(x-s) \hat{H}_\ell(x'-s) e^{-(x-s)^2} e^{-(x'-s)^2} dP_{S,n}(s).\end{aligned}$$

Denoting by $\Gamma_{\mathbf{x}}$ (resp. $\Gamma_{\mathbf{x}}^n$) the linear operators defined by (11), we consider the vanishing derivatives precertificates,

$$\eta_V^n = \sum_{i=1}^M \alpha_i^n (R_\varphi^n(\cdot, x_i) + \beta_i^n \partial_2 R_\varphi^n(\cdot, x_i)), \quad \eta_V = \sum_{i=1}^M \alpha_i (R_\varphi(\cdot, x_i) + \beta_i \partial_2 R_\varphi(\cdot, x_i)) \quad (64)$$

$$\text{where } \begin{pmatrix} \alpha^n \\ \beta^n \end{pmatrix} = (\Gamma_{\mathbf{x}}^{n*} \Gamma_{\mathbf{x}}^n)^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix} \quad (65)$$

By Proposition 3 above, $\eta_V(t) < 1$ for $t \in X \setminus \{x_i\}_{i=1}^M$, $\eta_V(x_i) = 1$, and $\eta_V''(x_i) < 0$ for all $1 \leq i \leq M$. We prove that the same holds for η_V^n by uniform convergence of the derivatives.

We note that as $n \rightarrow +\infty$

$$\|(\Gamma_{\mathbf{x}}^{n*} \Gamma_{\mathbf{x}}^n)^{-1} - (\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})^{-1}\| \leq \frac{\|\text{Id} - (\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})^{-1} (\Gamma_{\mathbf{x}}^{n*} \Gamma_{\mathbf{x}}^n)\|}{1 - \|\text{Id} - (\Gamma_{\mathbf{x}}^* \Gamma_{\mathbf{x}})^{-1} (\Gamma_{\mathbf{x}}^{n*} \Gamma_{\mathbf{x}}^n)\|} \rightarrow 0,$$

therefore $\alpha^n \rightarrow \alpha$ and $\beta^n \rightarrow \beta$. Moreover, (49) implies that for $k \in \{0, 1, 2\}$ and $\ell \in \{0, 1\}$,

$$\sup_{x \in X} |(\partial_1)^k(\partial_2)^\ell R_\varphi^n(x, x_i) - (\partial_1)^k(\partial_2)^\ell R_\varphi(x, x_i)| \rightarrow 0,$$

which implies that $\sup_{x \in X} |\eta_V^{n(k)}(x) - \eta_V^{(k)}(x)| \rightarrow 0$. Observing that there is some $r > 0$, some $\varepsilon > 0$, such that

$$\begin{aligned}\eta_V &\leq 1 - \varepsilon && \text{in } X \setminus \bigcup_{i=1}^M (x_i - r, x_i + r), \\ \eta_V'' &< 0 && \text{in } X \cap \left(\bigcup_{i=1}^M (x_i - r, x_i + r) \right), \\ \eta_V &< 1 && \text{in } X \cap \left(\bigcup_{i=1}^M (x_i - r, x_i + r) \setminus \{x_i\} \right).\end{aligned}$$

we conclude by standard arguments of uniform convergence that the same holds for η_V^n for n large enough. This ensures the Non-Degenerate Source Condition holds for m_0 with $P_{S,n}$.

As for the similar statement concerning the $2M - 1$ Non-Degeneracy of x_0 , we skip the proof as it is the same pattern, involving η_W and the matrices Ψ_{2M-1} instead. \square

A.4 Proof of Proposition 5

We begin with a continuous extension lemma for rescaled determinants.

Lemma 5. *Let $\{u_1, \dots, u_n\} \subset \mathcal{C}^{n-1}(\mathbb{R})$.*

Then the function

$$U : (s_1, \dots, s_n) \mapsto \frac{1}{\prod_{1 \leq i < j \leq n} (s_j - s_i)} \begin{vmatrix} u_1(s_1) & \cdots & u_n(s_1) \\ \vdots & & \vdots \\ u_1(s_n) & \cdots & u_n(s_n) \end{vmatrix} \quad (66)$$

has a continuous extension to \mathbb{R}^n . Moreover, there exists some constant $\gamma > 0$ such that

$$\forall t \in \mathbb{R}, \quad U(s, \dots, s) = \gamma \begin{vmatrix} u_1(s) & \cdots & u_n(s) \\ u'_1(s) & \cdots & u'_n(s) \\ \vdots & & \vdots \\ u_1^{(n-1)}(s) & \cdots & u_n^{(n-1)}(s) \end{vmatrix}. \quad (67)$$

Proof. We prove the continuity by operations on rows in the determinant. As we operate on rows only, we only represent the j -th column. Applying a Taylor expansion and subtracting the first row to its successors, then the second one to its successors,

$$\begin{aligned} \begin{vmatrix} u_j(s_1) \\ \vdots \\ u_j(s_n) \end{vmatrix}_{1 \leq j \leq n} &= \left(\prod_{i>2} (s_i - s_1) \right) \int_0^1 \begin{vmatrix} u'_j(s_1(1 - \theta_1) + s_2\theta_1) \\ \vdots \\ u'_j(s_1(1 - \theta_1) + s_n\theta_1) \end{vmatrix}_{1 \leq j \leq n} d\theta_1 \\ &= \left(\prod_{i>1} (s_i - s_1) \right) \left(\prod_{i>2} (s_i - s_2) \right) \\ &\quad \times \int_{[0,1]^2} \theta_1^{n-2} \begin{vmatrix} u'_j(s_1(1 - \theta_1) + s_2\theta_1) \\ u''_j(s_1(1 - \theta_1) + s_2\theta_1(1 - \theta_2) + s_3\theta_1\theta_2) \\ \vdots \\ u''_j(s_1(1 - \theta_1) + s_2\theta_1(1 - \theta_2) + s_n\theta_1\theta_2) \end{vmatrix}_{1 \leq j \leq n} d\theta_1 d\theta_2. \end{aligned}$$

Iterating this procedure, we see that we may write

$$\begin{vmatrix} u_j(s_1) \\ \vdots \\ u_j(s_n) \end{vmatrix}_{1 \leq j \leq n} = \left(\prod_{k>l} (s_k - s_l) \right) \int_{[0,1]^{n-1}} \theta_1^{\alpha_1} \cdots \theta_{n-2}^{\alpha_{n-2}} \quad (68)$$

$$\times \begin{vmatrix} u_j(s_1) \\ u'_j(s_1(1 - \theta_1) + s_2\theta_1) \\ u''_j(g_3(s_1, s_2, s_3, \theta_1, \theta_2)) \\ \vdots \\ u_j^{(n-1)}(g_{n-1}(s_1, \dots, s_n, \theta_1, \dots, \theta_{n-1})) \end{vmatrix}_{1 \leq j \leq n} d\theta_1 \dots d\theta_{n-1} \quad (69)$$

for some exponents $\alpha_1, \dots, \alpha_{n-2}$ in \mathbb{N} , and with

$$\begin{aligned} g_k(s_1, \dots, s_k, \theta_1, \dots, \theta_{k-1}) &= s_1(1 - \theta_1) + s_2\theta_1(1 - \theta_2) + \cdots \\ &\quad + s_{k-1}\theta_1\theta_2 \dots \theta_{k-2}(1 - \theta_{k-1}) + s_k\theta_1\theta_2 \dots \theta_{k-2}\theta_{k-1} \end{aligned}$$

which describes some convex combination of s_1, \dots, s_k , for $1 \leq k \leq n-1$.

This yields the claimed continuity and value at (s, \dots, s) . \square

Now, we are in position to prove Proposition 5.

Proof of Proposition 5. First, we note that $A_V(s_1, \dots, s_{2M}) > 0$ (resp. $A_W(s_1, \dots, s_{2M}) > 0$) always holds by the extended total positivity of the Gaussian kernel. Next, we show that $B_V(t, s_1, \dots, s_{2M}) > 0$ holds.

Now, we introduce a renormalized version of B_V ,

$$F(t, s_1, \dots, s_{2M}) = \frac{|t|^{2M}}{\prod (t - x_i)^2} \frac{1}{\prod_{i < j} (s_j - s_i)} \begin{vmatrix} 1 & \psi(t, s_1) & \cdots & \psi(t, s_{2M}) \\ 1 & \psi(x_1, s_1) & \cdots & \psi(x_1, s_{2M}) \\ 0 & \partial_1 \psi(x_1, s_1) & \cdots & \psi(x_1, s_{2M}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \psi(x_M, s_1) & \cdots & \psi(x_M, s_{2M}) \\ 0 & \partial_1 \psi(x_M, s_1) & \cdots & \psi(x_M, s_{2M}) \end{vmatrix} \quad (70)$$

Let $r' > 0$. By Lemma 5, which is a refined version of Lemma 3, we note that H is continuous on $\mathbb{R} \times [s_* - r', s_* + r']$. Now, we prove that it can be extended by continuity to the compact set $\overline{\mathbb{R}} \times [s_* - r', s_* + r']$, where $\overline{\mathbb{R}} = [-\infty, +\infty]$, equipped with a suitable metric¹.

From the proof of Lemma 5, we know that H may be written as

$$F(t, s_1, \dots, s_{2M}) = \frac{|t|^{2M}}{\prod (t - x_i)^2} \int_{[0,1]^{2M-1}} \theta_1^{\alpha_1} \dots \theta_{2M-2}^{\alpha_{2M-2}} H(t, (s_j)_{1 \leq j \leq 2M}, (\theta_j)_{1 \leq j \leq 2M-1}) d\theta_1 \dots d\theta_{2M-1}. \quad (71)$$

where H is the determinant of the matrix whose first row is:

$$(1 \quad \psi(t, s_1) \quad \partial_2 \psi(t, g_1((s_j)_{1 \leq j \leq 2}, \theta_1)) \quad \cdots \quad (\partial_2)^{2M-1} \psi(t, g_{2M-1}((s_j)_{1 \leq j \leq 2M-1}, (\theta_j)_{1 \leq j \leq 2M-2}))) \quad (72)$$

and the rows $1 + 2i$ and $2 + 2i$ are

$$\begin{pmatrix} 1 & \psi(x_i, s_1) & \partial_2 \psi(x_i, g_1((s_j)_{1 \leq j \leq 2}, \theta_1)) & \cdots & (\partial_2)^{2M-1} \psi(x_i, g_{2M-1}((s_j)_{1 \leq j \leq 2M-1}, (\theta_j)_{1 \leq j \leq 2M-2})) \\ 0 & \partial_1 \psi(x_i, s_1) & \partial_1 \partial_2 \psi(x_i, g_1((s_j)_{1 \leq j \leq 2}, \theta_1)) & \cdots & \partial_1 (\partial_2)^{2M-1} \psi(x_i, g_{2M-1}((s_j)_{1 \leq j \leq 2M-1}, (\theta_j)_{1 \leq j \leq 2M-2})) \end{pmatrix} \quad (73)$$

Here $g_k(s_1, \dots, s_k, \theta_1, \dots, \theta_{k-1})$ is a convex combination of s_1, \dots, s_k .

Since $\{s_j\}_{j=1}^{2M} \subset [s_* - r', s_* + r']$, we see that, as $|t| \rightarrow +\infty$, the first row (72) converges uniformly to the row $(1 \quad 0 \quad 0 \quad \cdots \quad 0)$. As a result, since $\frac{|t|^{2M}}{\prod (t - x_i)^2} \sim 1$ and the other rows (73) are bounded, we obtain that $F(t, s_1, \dots, s_{2M})$ converges uniformly in $(s_j)_{j=1}^{2M}$ towards

$$F(\pm\infty, s_1, \dots, s_{2M}) \stackrel{\text{def.}}{=} \int_{[0,1]^{2M-1}} \theta_1^{\alpha_1} \dots \theta_{2M-2}^{\alpha_{2M-2}} H(\pm\infty, (s_j)_{1 \leq j \leq 2M}, (\theta_j)_{1 \leq j \leq 2M-1}) d\theta_1 \dots d\theta_{2M-1}. \quad (74)$$

where $H(\pm\infty, (s_j)_{1 \leq j \leq 2M}, (\theta_j)_{1 \leq j \leq 2M-1})$ is the determinant of the matrix whose first row is $(1 \quad 0 \quad \cdots \quad 0)$ and the other rows are given by (73).

As a result F is continuous on the compact set $\overline{\mathbb{R}} \times [s_* - r', s_* + r']$, hence uniformly continuous. For all $\varepsilon > 0$, there exists $r > 0$ such that $\max_i |s_i - s_*| \leq r$ implies

$$|F(t, s_1, \dots, s_{2M}) - F(t, s_*, \dots, s_*)| \leq \varepsilon. \quad (75)$$

¹For instance with $d(t, t') = |\arctan(t) - \arctan(t')|$.

Now, there is some constant $\gamma > 0$ such that

$$F(t, s_*, \dots, s_*) = \frac{|t|^{2M}}{\prod (t - x_i)^2} \gamma \begin{vmatrix} 1 & \psi(t, s_*) & \partial_2 \psi(t, s_*) & \dots & (\partial_2)^{2M-1} \psi(t, s_*) \\ 1 & \psi(x_1, s_*) & \partial_2 \psi(x_1, s_*) & \dots & (\partial_2)^{2M-1} \psi(x_1, s_*) \\ 0 & \partial_1 \psi(x_1, s_*) & \partial_1 \partial_2 \psi(x_1, s_*) & \dots & \partial_1 (\partial_2)^{2M-1} \psi(x_1, s_*) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \psi(x_M, s_*) & \partial_2 \psi(x_M, s_*) & \dots & (\partial_2)^{2M-1} \psi(x_M, s_*) \\ 0 & \partial_1 \psi(x_M, s_*) & \partial_1 \partial_2 \psi(x_M, s_*) & \dots & \partial_1 (\partial_2)^{2M-1} \psi(x_M, s_*) \end{vmatrix}$$

We have already met similar determinants in the proof of Proposition 3. They characterize the number of roots (including multiplicity) of a function of the form

$$t \mapsto b_0 + (b_{1,0} + b_{1,1}t + \dots + b_{1,2M-1}t^{2M-1})e^{-(t-s_*)^2/2}, \quad (76)$$

which is at most $2M$ by [41, Lemma 2.7]. Therefore, arguing as in Proposition 3, it is possible to show that this determinant is positive, including for $t = x_i$. Moreover $F(\pm\infty, s_*, \dots, s_*) > 0$ by extended total positivity of the Gaussian. As a result, the function $t \mapsto F(t, s_*, \dots, s_*)$ has a lower bound $\ell > 0$ on $[-\infty, +\infty]$.

In (75) it is sufficient to choose $\varepsilon < \ell/2$, and for the corresponding $r > 0$, $\max_i |s_i - s_*| \leq r$ implies $F(t, s_1, \dots, s_{2M}) > 0$, hence $B_V(t, s_1, \dots, s_{2M}) > 0$ for $s_* - r \leq s_1 < \dots < s_{2M} \leq s_* + r$. We conclude by applying Proposition 1. The case of x_0 being $2M-1$ non-degenerate is similar. \square

References

- [1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford mathematical monographs. Oxford University Press, 2000.
- [2] Brett Bernstein and Carlos Fernandez-Granda. Deconvolution of point sources: a sampling theorem and robustness guarantees. *arXiv preprint arXiv:1707.00808*, 2017.
- [3] B.N. Bhaskar and B. Recht. Atomic norm denoising with applications to line spectral estimation. *In 2011 49th Annual Allerton Conference on Communication, Control, and Computing*, pages 261–268, September 2011.
- [4] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations: Theory, algorithms and performance bounds. *IEEE Signal Processing Magazine*, 25(2):31–40, 2008.
- [5] K. Bredies and H.K. Pikkarainen. Inverse problems in spaces of measures. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(1):190–218, 2013.
- [6] E.J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.
- [7] E.J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [8] Yuejie Chi, Louis L Scharf, Ali Pezeshki, and A Robert Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, 59(5):2182–2195, 2011.

- [9] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973.
- [10] Laurent Condat and Akira Hirabayashi. Cadzow Denoising Upgraded: A New Projection Method for the Recovery of Dirac Pulses from Noisy Linear Measurements. *Sampling Theory in Signal and Image Processing*, 14(1):17–47, 2015.
- [11] Y. de Castro and F. Gamboa. Exact reconstruction using beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336 – 354, 2012.
- [12] L. Demanet and N. Nguyen. The recoverability limit for superresolution via sparsity. *preprint arXiv:1502.01385*, 2015.
- [13] Q. Denoyelle, V. Duval, and G. Peyré. Support recovery for sparse super-resolution of positive measures. *to appear in Journal of Fourier Analysis and Applications*, 2017.
- [14] Q. Denoyelle, E. Soubies, V. Duval, and G. Peyré. Off-the-grid superresolution for multi-angle microscopy. *In preparation*, 2017.
- [15] D.L. Donoho. Super-resolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–1331, 9 1992.
- [16] C. Dossal and S. Mallat. Sparse spike deconvolution with minimum scale. In *Proceedings of SPARS*, pages 123–126, November 2005.
- [17] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
- [18] V. Duval and G. Peyré. Sparse spikes super-resolution on thin grids i: the lasso. *to appear in Inverse Problems*, 2017.
- [19] Carlos Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference: A Journal of the IMA*, 5(3):251–303, 2016.
- [20] J.J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.
- [21] M. Grasmair, O. Scherzer, and M. Haltmeier. Necessary and sufficient conditions for linear convergence of ℓ_1 -regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, 2011.
- [22] Hayit Greenspan. Super-Resolution in Medical Imaging. *The Computer Journal*, 52(1):43–63, January 2009.
- [23] Rémi Gribonval, Gilles Blanchard, Nicolas Keriven, and Yann Traonmilin. Compressive Statistical Learning with Random Feature Moments. Main novelties compared to version 1: improved concentration bounds, improved sketch sizes for compressive k-means and compressive GMM that now scale linearly with the ambient dimension, December 2017.
- [24] P. Heins. Reconstruction using local sparsity. *PhD thesis*, 2014.
- [25] Yingbo Hua and Tapan K Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(5):814–824, 1990.

- [26] Y. de Castro J.-M. Azaïs and F. Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
- [27] Thomas Kailath. ESPRIT—estimation of signal parameters via rotational invariance techniques. *Optical Engineering*, 29(4):296, 1990.
- [28] S. Karlin. *Total Positivity*. Number vol. 1 in Total Positivity. Stanford University Press, 1968.
- [29] S. Karlin and W.J. Studden. *Tchebycheff systems: with applications in analysis and statistics*. Pure and applied mathematics. Interscience Publishers, 1966.
- [30] V. Khaidukov, E. Landa, and T. Moser. Diffraction imaging by focusing/defocusing: An outlook on seismic superresolution. *GEOPHYSICS*, 69(6):1478–1490, June 2004.
- [31] S. Levy and P. Fullagar. Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution. *Geophysics*, 46:1235–1243, 1981.
- [32] Charles W. McCutchen. Superresolution in Microscopy and the Abbe Resolution Limit. *JOSA*, 57(10):1190–1192, October 1967.
- [33] V. I. Morgenshtern and E. J. Candès. Super-resolution of positive sources: the discrete setup. *preprint arXiv:1504.00717*, 2014.
- [34] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Transactions on Antennas and Propagation*, 42(10):1386–1391, October 1994.
- [35] J. W. Odendaal, E. Barnard, and C. W. I. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *IEEE Transactions on Antennas and Propagation*, 42(10):1386–1391, October 1994.
- [36] Clarice Poon and Gabriel Peyr. Multi-dimensional Sparse Super-resolution. *arXiv:1709.03157 [math]*, September 2017. arXiv: 1709.03157.
- [37] KG Puschmann and F Kneer. On super-resolution in astronomical imaging. *Astronomy & Astrophysics*, 436(1):373–378, 2005.
- [38] George Plya and Gabor Szeg. *Problems and Theorems in Analysis I*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1972. DOI: 10.1007/978-3-642-61983-0.
- [39] F. Santosa and W.W. Symes. Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1307–1330, 1986.
- [40] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop on*, pages 45–48. IEEE, 2015.
- [41] Geoffrey Schiebinger, Elina Robeva, and Benjamin Recht. Superresolution without separation. *to appear in Information and Inference: a Journal of the IMA*, 2015.
- [42] D.L. Donoho S.S. Chen and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

- [43] Gongguo Tang. Resolution limits for atomic decompositions via markov-bernstein type inequalities. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 548–552. IEEE, 2015.
- [44] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Sparse recovery over continuous dictionaries-just discretize. In *Signals, Systems and Computers, 2013 Asilomar Conference on*, pages 1043–1047. IEEE, 2013.
- [45] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE transactions on information theory*, 59(11):7465–7490, 2013.
- [46] Gongguo Tang and Benjamin Recht. Atomic decomposition of mixtures of translation-invariant signals. *IEEE CAMSAP*, 2013.
- [47] R. Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.
- [48] Yann Traonmilin, Nicolas Keriven, Rémi Gribonval, and Gilles Blanchard. Spikes super-resolution with random Fourier sampling. SPARS workshop 2017, June 2017.
- [49] J.A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.
- [50] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer Berlin Heidelberg, 1995.